How do new transit stations affect people's sentiment and activity? A case study based on social media data in Hong Kong

Chang, Haoliang; Huang, Jianxiang; Yao, Weiran; Zhao, Weizun; Li, Lishuai

DOI
10.1016/j.tranpol.2022.03.011

Publication date
2022

Document Version
Final published version

Published in
Transport Policy

Transport Policy 120 (2022) 139–155

Contents lists available at ScienceDirect

Transport Policy

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

How do new transit stations affect people's sentiment and activity? A case study based on social media data in Hong Kong

Haoliang Chang [a], Jianxiang Huang [b], Weiran Yao [c], Weizun Zhao [d], Lishuai Li [d,e,*]

[a] Department of Advanced Design and Systems Engineering, City University of Hong Kong, Hong Kong, China
[b] Department of Urban Planning and Design, The University of Hong Kong, Hong Kong, China
[c] Department of Civil and Environmental Engineering, Carnegie Mellon University, Pittsburgh, PA, 15213, USA
[d] School of Data Science, City University of Hong Kong, Hong Kong, China
[e] Faculty of Aerospace Engineering, TU Delft, Delft, the Netherlands

ARTICLE INFO

Keywords:
Tweet sentiment
Tweet activity
Urban rail transit
Twitter
Data mining
Social media data analysis

ABSTRACT

Urban rail development can increase land value, reduce commute time, and increase accessibility, as reported in the literature. However, little is known about the impact of opening urban rail transit stations on people's sentiment, particularly in the context of large metropolises where population density is significantly high. This paper investigates such impact by studying six new transit stations opened in Hong Kong. People's sentiment and activity in station nearby areas are estimated by tweet sentiment and tweet activity. We use the difference-in-difference model to study the impact of opening new transit stations. Tweet sentiment, tweet activity, tweet content, and footprints of people who visit the station-influenced area 'before and after' the opening of transit stations are analyzed. The results suggest that, in general, the introduction of transit stations causes a positive change in tweet activity, and the change is statistically significant after six months. Regarding tweet sentiment, new transit stations tend to pose a mixed effect in a short-term, a positive influence on areas with high-density residential places, yet a negative influence on areas with a large proportion of nature reserve areas. These short-term effects, positive or negative, become not significant in the long term (after twelve months). Our analysis also confirmed that the introduction of new transit stations increased accessibility from (to) other parts of the city to (from) the station's nearby area, which was shown by the expanded locations sustaining users visited. These findings indicate that the urban rail transit system in Hong Kong promotes more active neighborhoods yet does not always promotes positive influence on people's sentiment. Further studies are needed to make future urban rail transit systems promoting active and happy neighborhoods. The study is relevant to the Belt and Road Initiative (BRI) in methodologies, data, and findings. The social media analysis method used in this study, including text mining and sentiment analysis, can be easily extended to multiple language analysis for Singapore, Malaysia, as well as other regions in the belt and road plan. The developed tools could contribute to analyzing the influence of cross-country projects on local neighborhoods in the belt and road plan.

1. Introduction

Urban rail transit systems have been employed to improve mobility in cities worldwide. In some large high-density cities, such as Hong Kong, the transit system is well developed and continuously expanding to more neighborhoods and districts. Previous studies have been conducted to support and guide such development, focusing on the impact of mass transit on factors including economic activity (Canales et al., 2019; Credit, 2018; Dubé et al., 2014; Yu et al., 2018), transit ridership (Lee et al., 2013; Ryan and Frank, 2015), and environmental sustainability (Cervero and Sullivan, 2011; Loo and du Verle, 2017). However, the experience of city users who live, work or visit the urban transit stations' influenced area has been overlooked. Since intangible human life aspects, such as happiness, are gradually becoming important evaluation factors of successful urban planning (Krefis et al., 2018).

Most studies about rail transit system impact on urban form and people's life are related to the concept of transit-oriented development (TOD), which is an urban development type that maximizes residential, business, and leisure space within walking distance from public transport. Traditionally, TOD focused on suburban areas and greenfield sites

* Corresponding author. Present affiliation: Faculty of Aerospace Engineering, TU Delft, Delft, the Netherlands.
E-mail address: lishuai.li@tudelft.nl (L. Li).

https://doi.org/10.1016/j.tranpol.2022.03.011
Received 10 March 2022; Accepted 13 March 2022
Available online 21 March 2022

Fig. 1. Hong Kong's Mass Transit Railway (MTR) system map as of 2016 (Wikimedia Commons contributors, 2020).

where new towns or communities were being developed together with the transit system. Currently, TOD also includes urban development in large metropolitan cities that present well-developed transit systems already (Cervero and Murakami, 2009; Lee et al., 2013; Loo et al., 2010). Moreover, TOD is a core concept in recent transportation and urban planning studies (Brown and Werner, 2008; Nilsson and Delmelle, 2018). The five "Ds", which denote density, land use diversity, pedestrian-oriented design of neighborhoods, distance to transit, and destination accessibility, are commonly considered as the planning principles of TOD (Cervero and Murakami, 2008, 2009). Classic urban design principles, such as those proposed by Jacobs (1961), are also often used in TOD planning to cultivate community and vibrant city life.

The impact of transit access on property values has been well documented (Bowes and Ihlanfeldt, 2001; Cheshire and Sheppard, 1995; Debrezion et al., 2011; Duncan, 2011; Hess and Almeida, 2007; Seo et al., 2014); the benefits of TOD on transit efficiency, ridership, traffic congestion, and carbon emissions have also been studied extensively (Cervero and Sullivan, 2011; Freilich, 1998; Loo and du Verle, 2017; Zhang, 2010). Furthermore, some studies aimed to identify natural groupings in TODs with similar characteristics (Cervero and Murakami, 2009; Kamruzzaman et al., 2014; Loo and du Verle, 2017).

A few studies examined the impact of TOD on city users' living experience; some researchers argued that gentrification could be a result of TOD and may cause disruptions in low-income communities (Jones and Ley, 2016; Rayle, 2015; Soursourian, 2010). Other studies linked the development around transit access with pedestrian congestion (Lam et al., 1999; Ryan and Frank, 2015; Wang et al., 2013), pedestrian safety (González et al., 2019), traffic noise increase (Lam et al., 2009), air quality (Borrego et al., 2006), and the loss of open space (National Academies of Sciences, Engineering, and Medicine, 2005). However, whether the opening of new urban transit stations influences people's

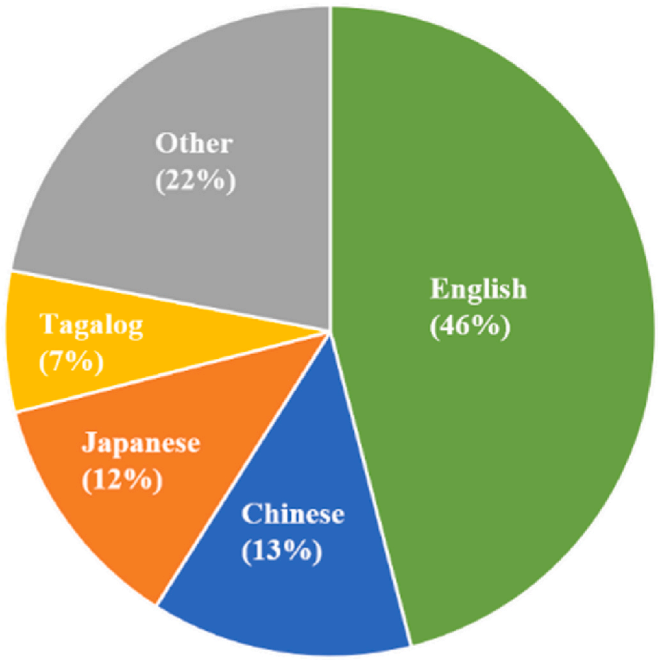

Fig. 2. Language distribution of the collected Twitter data.

happiness in surrounding neighborhoods or not remains unclear.

Meanwhile, emotional well-being[1] has received significant research interest in the field of economics, psychology, and other social science fields (Campbell et al., 1976; Graham, 2011; Helliwell and Putnam, 2004; Layard et al., 2014; McGillivray and Clarke, 2006; Oswald and Powdthavee, 2008). To measure emotional well-being and determine the factors that contribute to it, Layard et al. (2014) developed a state-of-the-art predictive model of well-being. Moreover, some studies focused on the effect of housing and residential conditions on people's happiness (Lu, 1999; Mohan and Twigg, 2007; Parkes et al., 2002; Permentier et al., 2011). However, only one study has been conducted to examine people's happiness near urban transit stations, which evaluated the effects of better rail access on homeowners' happiness using survey data (Wu, 2014).

Traditionally, researchers relied on self-reported subjective happiness and life satisfaction; however, the data collection process is costly, difficult to perform for large samples (Csikszentmihalyi and Larson, 2014), and controversial in reliability (Andrews and Withey, 2012; Krueger and Schkade, 2008). Currently, the rise of social media analytics provides new opportunities to study the relationship between people and their living environment. Despite sample bias, geocoded tweets have been found to be a useful tool for urban research (Hamstead et al., 2018; Lansley and Longley, 2016; Lloyd and Cheshire, 2017; Longley et al., 2015; Shao et al., 2017). Some studies focused on the spatial or temporal patterns of tweet activity and its relationship with urban forms and functions (Longley et al., 2015; Nadai et al., 2016). Some other studies employed text mining to analyze the content of social media data (Bertrand et al., 2013; Frank et al., 2013; Huang et al., 2016; Lansley and Longley, 2016; Quercia et al., 2012). Correlations between social media sentiment and the built environment have been found; strong sentiments were found near parks, transport hubs, and polluted areas in New York City (Bertrand et al., 2013). Outside the urban research field, many studies used Twitter data to investigate large-scale social problems, including happiness on a national level or large social networks (Dodds et al., 2011; Durahim and Coşkun, 2015; Frank et al., 2013), epidemic diseases (Fung et al., 2014; Nagar et al., 2014), election results (Bollen et al., 2011), terrorist attacks (Gruebner et al., 2016), etc.

The objective of this research is to gain a better understanding of how the urban transit system influences people's activity and sentiment as reflected on social media, in order to build more people-centric urban transit systems in the future. In this paper, tweet analysis is used to estimate people's social media sentiment and activity near urban transit stations. A quasi-experiment is then adopted to estimate the causal impact of the opening of new urban transit stations on tweet sentiment and activity in nearby neighborhoods. Moreover, we select the social media users who ever posted messages before and after the opening of transit stations as the station influenced users. We show how new transit stations change their daily lives by analyzing and comparing their tweet sentiment, tweet content, and tweet footprints before and after the introduction of a transit station.

The study is relevant to the "Belt and Road Initiative" (BRI) in methodologies, data, and findings. The data analysis methods developed from this study can serve as a novel assessment tool to evaluate the degree of success of BRI projects, a large proportion of which being urban rail infrastructure and TOD. The novelty of our methods allows for comprehensive monitoring of the emotional and behavioral impact of a BRI project, which is efficient and statistically more powerful, which can supplement, if not in replacement, of traditional survey-based approaches. Second, the choice of social media platform, Twitter, is relevant for key BRI countries and cities. Twitter has high penetration rates outside of mainland China, compared with alternative platforms such as Weibo. Lastly, findings from our study sites in Hong Kong are considered transferable to the key belt and road countries. Hong Kong is perceived globally as a success story of the compact and transit-driven city. Findings from the Hong Kong study sites are sentinel to TOD projects in other cities within the BRI policy priorities.

This paper is organized as follows. Section 2 introduces the study area, the general information about the social media data, and the census data we collected. Section 3 illustrates the methodology used to gauge the influence of the newly built transit stations on tweet sentiment and tweet activity; sentiment analysis and tweet text content analysis methods are also presented. Section 4 and Section 5 presents the results and discussions. Finally, Section 6 gives the conclusions.

2. Study area and data used

2.1. Study area: Hong Kong

Fig. 1 presents an overview of the Hong Kong Mass Transit Railway (MTR) service system. Hong Kong is an appropriate city for this study. Firstly, Hong Kong has been recognized as a successful case of rail transit investment and urban development integration (Cervero and Murakami, 2009); the city's significantly high urban density and intensive railway service render it a suitable place for this study. Moreover, the city utilizes limited land resources by promoting intensified development above or near railway stations. Secondly, public transport, particularly rail transit, plays a significant role in almost everyone's life in Hong Kong: more than 90% of all motorized trips in Hong Kong are performed by public transport, which is the highest figure in the world (Lam and Bell, 2002). Moreover, 5.46 million people used the MTR system each weekday by the end of 2014 (MTR, 2015), accounting for 77.2% of the city's population (Census and Statistics Department, 2011). Thirdly, MTR has applied the "Rail + Property" development program model extensively (Cervero and Murakami, 2009); buildings surround most of the metro stations, amounting to 13 million square meters of floor area (MTR, 2015). Based on this strategy, the Hong Kong government intensifies the relationship between urban transit stations and people, which also renders Hong Kong an appropriate city to study the relationship between the urban transit system and people.

[1] In this paper, we use the words "happiness" and "well-being" interchangeably, because the concept of subjective well-being (SWB) is used in the literature as a substitute for the term "happiness" (Diener et al., 2003).

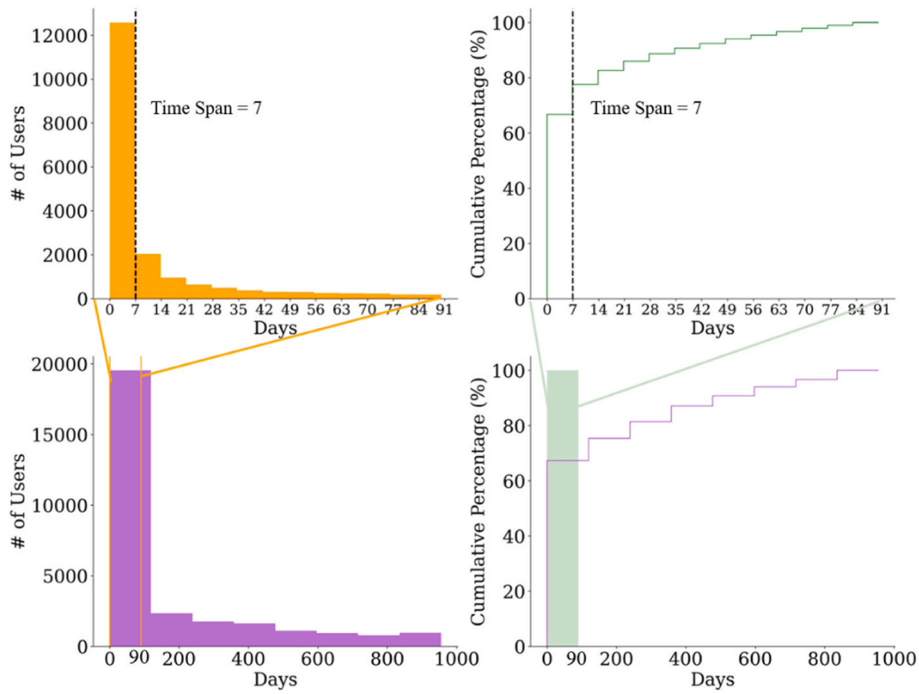

Fig. 3. Tweet time span distribution of Twitter users who are not bot and verified accounts.

Table 1
Tweet filtering process.

	Number of tweets	Number of unique users
Raw dataset general information		
Whole dataset	4,206,921	249,912
Filtering process		
Keep geocoded tweets	1,201,270	103,400
Use *ArcMap* to find tweets posted in Hong Kong	887,429	83,634
Consider tweets in Chinese and English only	590,143	63,459
Remove verified accounts	562,578	62,820
Delete bots	409,626	29,023
Remove Temporary Visitors	341,775	15,984
Dataset used		
Filtered dataset	341,775	15,984

2.2. Social media data

Hong Kong has high internet penetration (79.5%), smartphone penetration (85.8%), and mobile internet usage (96.5%) rates (Census and Statistics Department, 2016a). Popular social media platforms in Hong Kong include Facebook (85% penetration rate), YouTube (83%), WhatsApp (82%), Instagram (57%), WeChat (53%), Facebook Messenger (50%), LINE (29%), Twitter (26%), and Skype (24%) (Statista, 2019). Although Twitter does not have the highest penetration rate, Twitter is the biggest data source that offers open social media data in Hong Kong. Twitter releases an easy-to-use streaming API, which allows researchers to download the tweets posted in one specific study area. Other studies have shown that Twitter has gradually become a central site where people share their views with others, which makes it a great platform to study public opinions (Qi et al., 2020). Therefore, we chose Twitter as the data source to estimate people's sentiment and activity on social media.

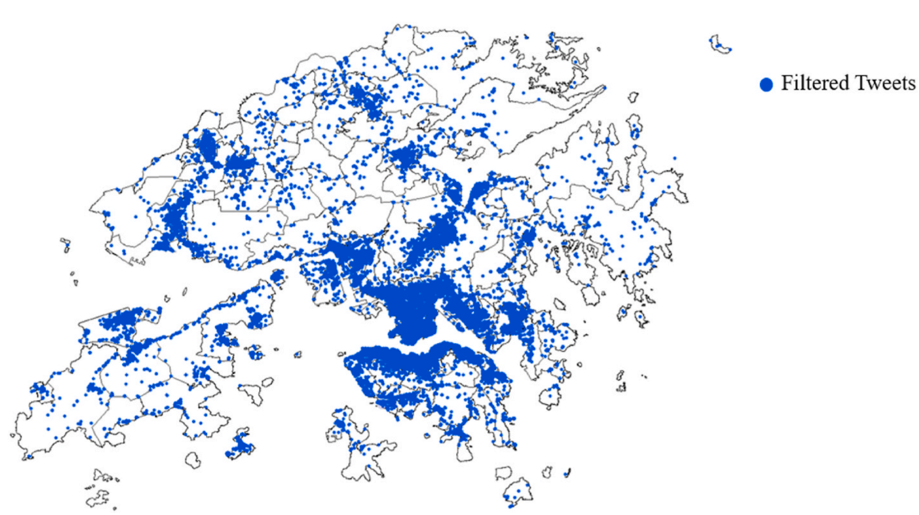

Fig. 4. Filtered tweets posted in Hong Kong from May 2016 to Dec 2018.

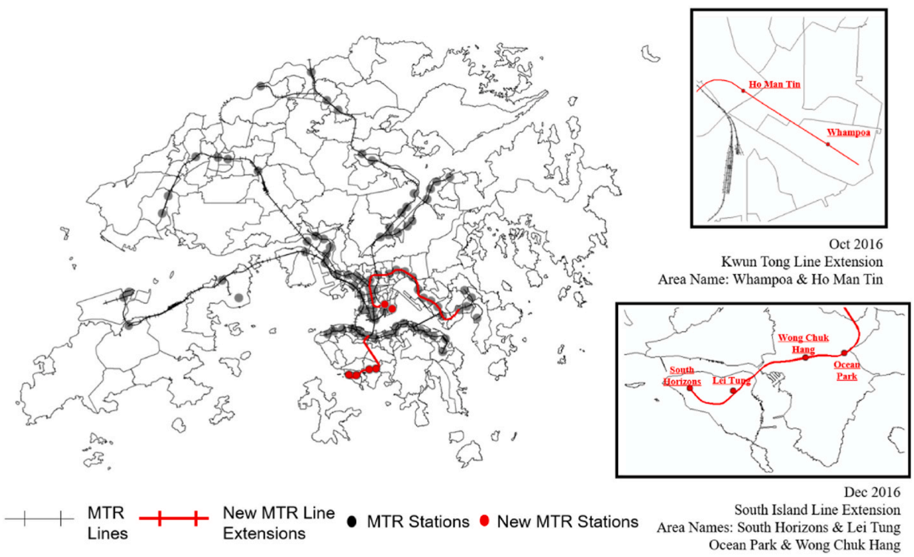

Fig. 5. Hong Kong's Mass Transit Railway line extensions and influenced areas.

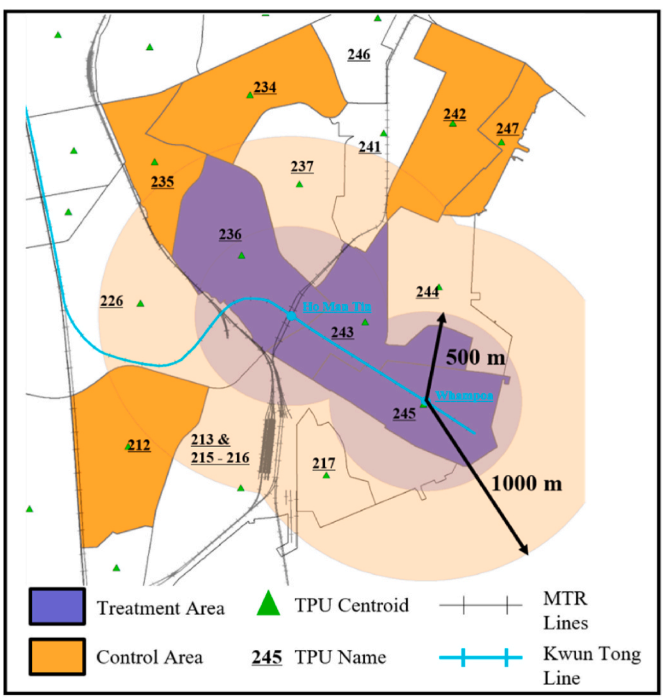

Fig. 6. Treatment area and control area for Whampoa & Ho Man Tin.

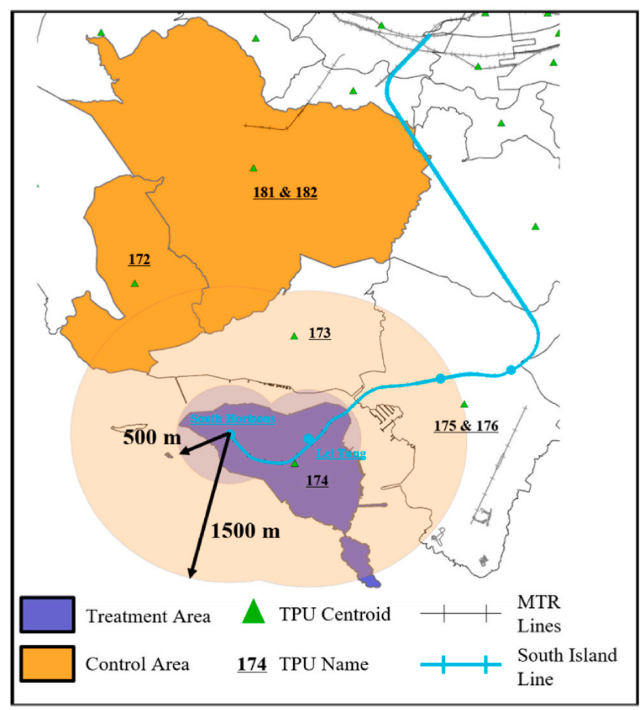

Fig. 7. Treatment area and control area for South Horizons & Lei Tung.

We constructed our Twitter database using Twitter's streaming API and collected 4,206,921 tweets posted from May 7, 2016 to December 18, 2018. The collected Twitter data included time, tweet text, user ID, language, *etc*. Some tweets are also tagged with accurate latitude and longitude information. Fig. 2 shows the language distribution of the collected tweets. English and Chinese tweets account for nearly 60% of all collected tweets.

After data collection, we filtered the tweets by keeping the geotagged tweets and remove tweets from bot accounts. Then we kept the Chinese tweets and English tweets in this study. At last, we removed tweets from temporary visitors. We computed the time span between the first tweet and the last tweet of each user after removing the bot and verified accounts. Fig. 3 shows the distribution of the time span. Users with a time span less than or equal to 7 days are likely to be temporary visitors. In this study, we removed the tweets posted by temporary visitors and only considered the other Twitter users who were likely to stay in Hong Kong for a longer time. The tweet filtering process is detailed in the appendix.

Finally, we obtained 341,775 geocoded tweets posted by 15,984 unique Twitter users (Table 1). Fig. 4 shows the locations of the filtered geocoded tweets collected from May 2016 to December 2018.

2.3. Census, land use and point of interest (POI) information

Factors such as population, income, and local land use of an area might influence tweet sentiment and tweet activity. Hence, we extracted population and median monthly income data from the 2016 census data collected by the Census and Statistics Department of Hong Kong (Census and Statistics Department, 2016b), and collected the land use from the

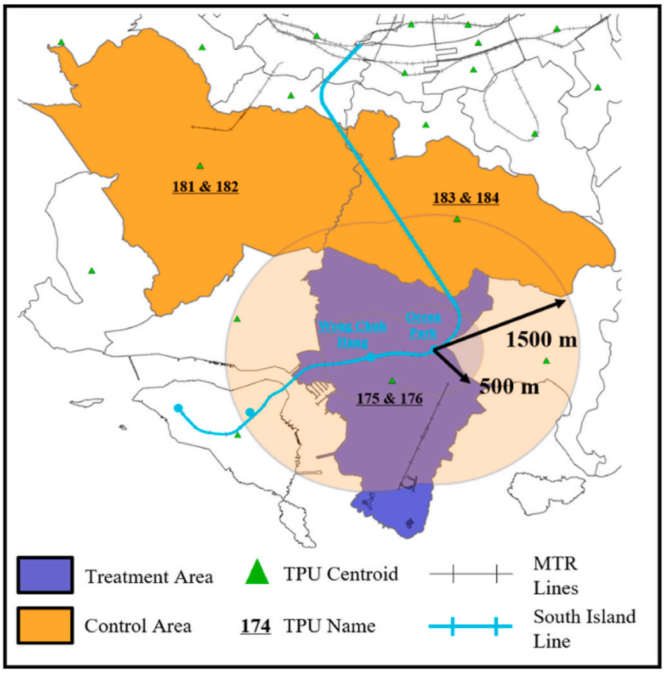

Fig. 8. Treatment area and control area for Ocean Park & Wong Chuk Hang.

Table 2
Twitter data collected for each influenced area.

Study Areas	Area type	Number of tweets	Number of unique users
Whampoa & Ho Man Tin	Treatment group	2356	648
	Control group	6231	2487
South Horizons & Lei Tung	Treatment group	948	322
	Control group	6295	3351
Ocean Park & Wong Chuk Hang	Treatment group	3988	1625
	Control group	6248	3341

Table 3
Population and median monthly income of each study area.

Study Areas	Area type	Number of TPUs	2016 Population	2016 Weighted Average of Median Monthly Income (HK$)
Whampoa & Ho Man Tin	Treatment group	3	111,454	16,691.64
	Control group	5	109,764	19,549.64
South Horizons & Lei Tung	Treatment group	1	86,752	17,500.00
	Control group	2	61,367	17,426.68
Ocean Park & Wong Chuk Hang	Treatment group	1	14,465	18,750.00
	Control group	2	9613	62,433.68

Hong Kong Land Department (Land Department, 2016) and the Point-of-interest (POI) data from OpenStreetMap (OpenStreetMap Contributors, 2017).

3. Methodology

3.1. Experiment design

The objective of this study is to gain a better understanding of how the opening of urban transit stations influences people's activity and sentiment as reflected on social media. A quasi-experiment is adopted to estimate the causal impact of the opening of new urban transit stations on tweet sentiment and activity in nearby neighborhoods. We selected three areas with new MTR transit stations and assessed nearby area's sentiment and activity change after the introduction of these stations. In 2016, two new line extensions—Kwun Tong Line Extension and South Island Line Extension—and six new MTR stations—Whampoa, Ho Man Tin, South Horizons, Lei Tung, Ocean Park, and Wong Chuk Hang—were opened. The location of all MTR stations, the studied line extensions, and the names of the influenced areas are presented in Fig. 5. The whole Hong Kong has been divided into many Tertiary Planning Units (TPUs). The TPU is the census tract used by Hong Kong's Census and Statistics Department to collect socio-demographic information (Census and Statistics Department, 2011, 2016b).

The sentiment level of the study areas was evaluated by subtracting the percentage of negative tweets from that of positive tweets, as performed by Durahim and Coşkun (2015). The number of tweets posted in a study area, called tweet activity, is used to estimate how many people would like to visit this area. Previous studies have demonstrated that it is appropriate to estimate people's mobility using Twitter data, including park visitation (Hamstead et al., 2018) and the mobility pattern analysis (Hawelka et al., 2014) as examples.

Difference-in-difference (DID) analysis was used to determine the influence of MTR transit station introduction on tweet sentiment and tweet activity. The treatment group and control group were selected carefully, as the prerequisite for DID analysis is the parallel trend assumption between the treatment group and control group before the introduction of the MTR stations. In this study, we employed the following strategy to define the treatment group and control group of each influenced area:

1. Treatment group: a TPU is assigned as the treatment group if the TPU′ centroid falls within a 500-m radius circle of a MTR station (Baker and Lee, 2019; Bardaka et al., 2018), as shown in Figs. 6–8. The 500-m criterion is commonly used in related studies, such as in Cervero and Murakami (2009).
2. Control group: a group of TPUs that intersect with a 1000-m (or 1500-m) radius circle around a MTR station but do not intersect with the 500-m circles derived from six new transit stations. The range, 1000- or 1500-m, was set based on which setting meets the parallel trend assumption better.

Based on the above criteria, the treatment group and control group for Whampoa & Ho Man Tin are shown in Fig. 6, the ones for South Horizons & Lei Tung are shown in Fig. 7, and those for Ocean Park & Wong Chuk Hang in Fig. 8.

The number of Twitter messages and unique Twitter user IDs for the treatment and control groups of three influenced areas are shown in Table 2.

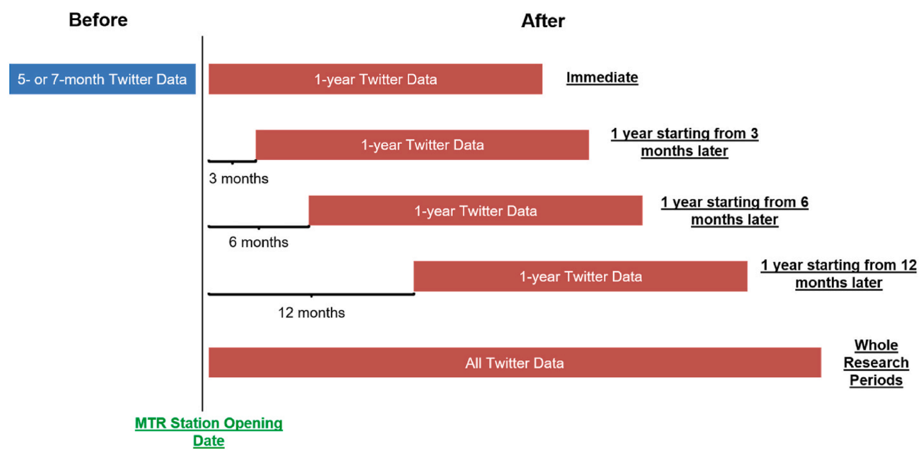

Fig. 9. DID analysis time settings.

Table 4
Studied time periods for the three influenced areas.

Study Areas	Condition	Analysis time span				
		Immediate	3-month period	6-month period	12-month period	Whole research periods
Whampoa & Ho Man Tin	Before	May 2016 – Sep 2016	May 2016 – Sep 2016	May 2016 – Sep 2016	May 2016 – Sep 2016	May 2016 – Sep 2016
	After	Nov 2016 – Oct 2017	Feb 2017 – Jan 2018	May 2017–Apr 2018	Nov 2017–Oct 2018	Nov 2016–Dec 2018
South Horizons & Lei Tung; Ocean Park & Wong Chuk Hang	Before	May 2016 – Nov 2016	May 2016 – Nov 2016	May 2016 – Nov 2016	May 2016 – Nov 2016	May 2016 – Nov 2016
	After	Jan 2017–Dec 2017	Apr 2017–Mar 2018	July 2017–June 2018	Jan 2018–Dec 2018	Jan 2017–Dec 2018

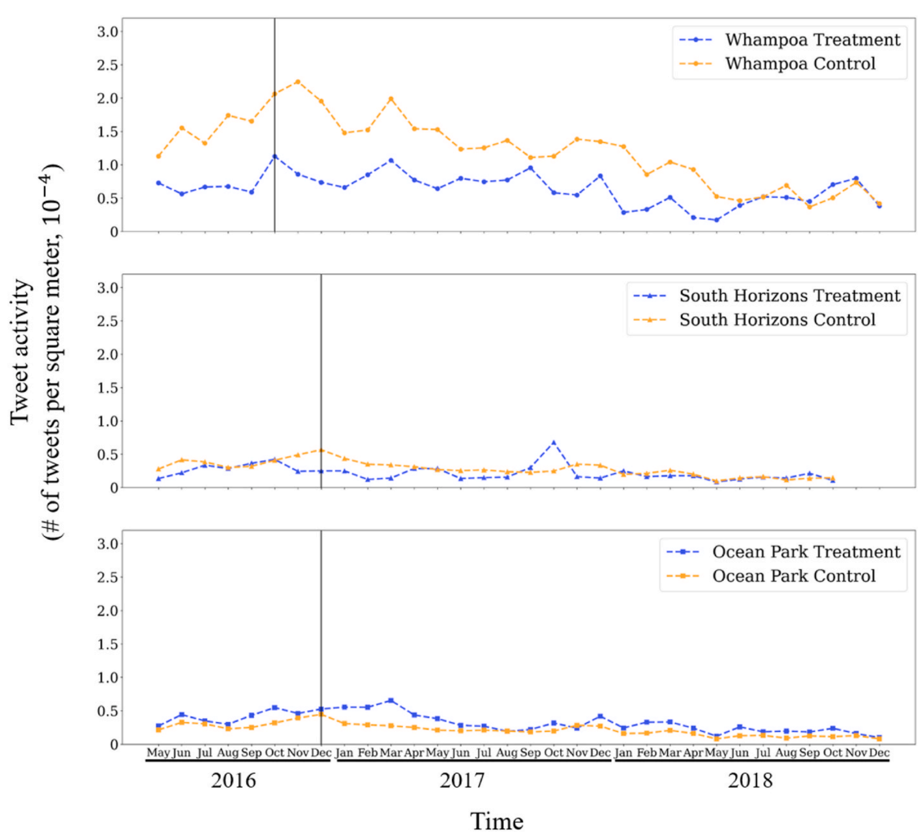

Fig. 10. Tweet activity change over time for three MTR station-influenced areas.

Table 5
Overall tweet activity difference-in-difference analysis results.

	Immediate	3-month period	6-month period	12-month period	Whole research period
$T_i \times POST_{i,t}$	0.1447 (0.125)	0.0865 (0.122)	0.0470 (0.131)	0.2400* (0.137)	0.1731 (0.155)
T_i	−1.2932*** (0.137)	−1.2713*** (0.131)	−1.3256*** (0.141)	−1.3086*** (0.148)	−1.2928*** (0.169)
$POST_{i,t}$	−0.1447 (0.088)	−0.1825** (0.086)	−0.2681*** (0.093)	−0.7352*** (0.097)	−0.4192*** (0.110)
$ln(Pop_i)$	−2.1322*** (0.404)	−1.8164*** (0.389)	−1.6016*** (0.419)	−1.7662*** (0.440)	−2.1021*** (0.422)
$ln(Income_i)$	−1.3287*** (0.130)	−1.1650*** (0.125)	−1.1631*** (0.135)	−1.2190*** (0.142)	−1.3106*** (0.136)
Area Fixed Effect	Yes	Yes	Yes	Yes	Yes
No. of observations	96	90	90	90	168
AIC	43.20	28.49	41.99	50.73	179.28
Adjusted R-squared	0.858	0.867	0.854	0.846	0.754

Notes: *, **, and *** indicate statistical significances of 0.1, 0.05, and 0.01, respectively. Standard errors are presented in parentheses. AIC stands for Akaike Information Criterion. The area fixed effect has been added to the model.

Table 6
Tweet activity difference-in-difference analysis results for three MTR station-influenced areas.

Influenced area		Immediate	3-month period	6-month period	12-month period	Whole research period
Whampoa & Ho Man Tin	$T_i \times POST_{i,t}$	0.1678 (0.141)	0.1545 (0.181)	0.0492 (0.237)	0.2303 (0.305)	0.2466 (0.317)
	T_i	−1.1517*** (0.118)	−1.1517*** (0.152)	−1.1517*** (0.199)	−1.1517*** (0.256)	−1.1517*** (0.290)
	$POST_{i,t}$	0.0212 (0.099)	−0.0625 (0.128)	−0.2067 (0.168)	−0.6673*** (0.216)	−0.3727 (0.224)
	Num	34	34	34	34	62
	AIC	−13.83	3.36	21.73	38.81	83.16
	Adjusted R-squared	0.888	0.827	0.763	0.650	0.519
South Horizons & Lei Tung	$T_i \times POST_{i,t}$	−0.0746 (0.239)	0.0985 (0.226)	0.1211 (0.266)	0.2454 (0.213)	0.0708 (0.243)
	T_i	−1.9339*** (0.190)	−1.9339*** (0.180)	−1.9339*** (0.212)	−1.9339*** (0.163)	−1.9339*** (0.211)
	$POST_{i,t}$	−0.2061 (0.169)	−0.3354** (0.160)	−0.5021** (0.188)	−0.8115*** (0.151)	−0.4813*** (0.172)
	Num	38	38	38	34	58
	AIC	33.11	28.79	41.20	19.65	60.84
	Adjusted R-squared	0.889	0.890	0.854	0.909	0.856
Ocean Park & Wong Chuk Hang	$T_i \times POST_{i,t}$	0.0842 (0.192)	0.0412 (0.152)	0.0425 (0.199)	0.1669 (0.203)	0.1255 (0.244)
	T_i	−0.5463*** (0.152)	−0.5463*** (0.121)	−0.5463*** (0.158)	−0.5463*** (0.161)	−0.5463** (0.214)
	$POST_{i,t}$	−0.1890 (0.136)	−0.3140*** (0.107)	−0.4685*** (0.141)	−0.8158*** (0.144)	−0.5024*** (0.172)
	Num	38	38	38	38	62
	AIC	16.27	−1.49	19.17	20.55	66.56
	Adjusted R-squared	0.430	0.628	0.556	0.654	0.327

We have also included two social-demographic control variables that may affect twitter activity and sentiment: population and income. The population and median monthly income of the treatment group and control group of each study area are summarized in Table 3. Here we estimated the median monthly income of each TPU group by computing the weighted average of the median monthly income of TPUs in this TPU group using each TPU's population as the weight.

To determine whether the introduction of new MTR transit stations significantly changes tweet sentiment and tweet activity in nearby areas or not, we built the following DID models for the study areas using ordinary least squares (OLS) regression:

$$Senti_{i,t} = \beta_0 + \beta_1 T_i + \beta_2 POST_{i,t} + \beta_3 T_i \times POST_{i,t} + \beta_4 ln(Pop_i) + \beta_5 ln(Income_i) + Area_i + \varepsilon_{i,t} \quad (1)$$

$$ln(Activity_{i,t}) = \beta_0 + \beta_1 T_i + \beta_2 POST_{i,t} + \beta_3 T_i \times POST_{i,t} + \beta_4 ln(Pop_i) + \beta_5 ln(Income_i) + Area_i + \varepsilon_{i,t} \quad (2)$$

where:

- i denotes a TPU group (treatment group or control group of a study area) and t represents month.
- $Senti_{i,t}$ represents the sentiment of tweets posted in the TPU group i in month t, while $ln(Activity_{i,t})$ is the natural log of the tweet activity of the TPU group i in month t. Since the DID analysis was conducted monthly, $Senti_{i,t}$ was computed as the percentage of positive tweets minus the percentage of negative tweets in TPU group i in month t; $Activity_{i,t}$ was computed for the number of tweets posted in TPU group i in month t.
- $POST_{i,t}$ and T_i are binary indicator variables. $POST_{i,t} = 1$ represents after the opening of the corresponding MTR stations; $POST_{i,t} = 0$ represents before. T_i is an indicator for the TPU group i being the treatment group or not; T_i equals to one if treatment group, and zero if control group.
- $ln(Pop_i)$ represents the natural log of the total population of the TPU group i. $ln(Income_i)$ denotes the natural log of the approximated median monthly income of the TPU group i. We approximated the median monthly income of each TPU group by computing the weighted average of the median monthly income of TPUs in this TPU group using each TPU's population as weight.
- $Area_i$ denotes the study area fixed effect (Whampoa & Ho Man Tin, South Horizons & Lei Tung, or Ocean Park & Wong Chuk Hang).
- β_3 is a difference-in-difference estimator given by the integration of T_i and $POST_{i,t}$ and is the coefficient of interest. It measures the tweet sentiment and tweet activity effect owing to the introduction of the MTR stations. If this coefficient is statistically significant in the DID models, the changes in the variables of interest of the treatment area are significantly different from those of the control area after line

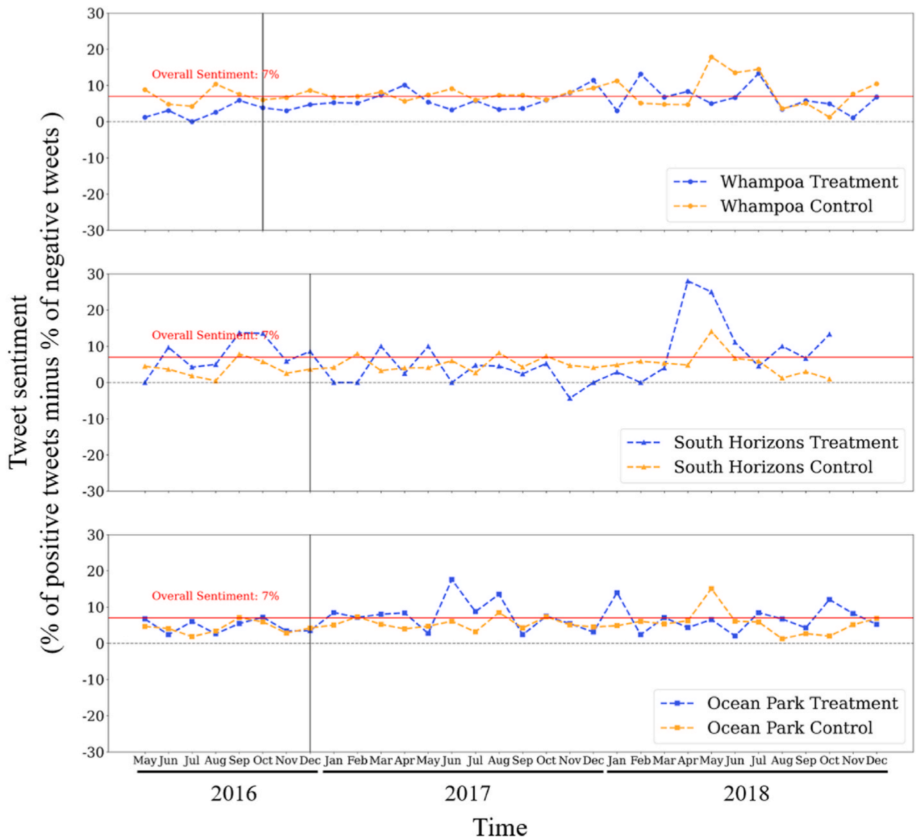

Fig. 11. Tweet sentiment change over time for three MTR station-influenced areas.

Table 7
Overall tweet sentiment difference-in-difference analysis results.

	Immediate	3-month period	6-month period	12-month period	Whole research period
$T_i \times POST_{i,t}$	0.0030 (0.012)	0.0014 (0.015)	0.0071 (0.018)	0.0224 (0.021)	0.0091 (0.016)
T_i	−0.0313** (0.014)	−0.0321** (0.016)	−0.0214 (0.019)	−0.0346 (0.023)	−0.0265 (0.017)
$POST_{i,t}$	0.0108 (0.009)	0.0099 (0.01)	0.0090 (0.013)	0.0132 (0.015)	0.0126 (0.011)
$ln(Pop_i)$	0.0945** (0.04)	0.0749 (0.047)	0.0668 (0.057)	0.1856*** (0.067)	0.1020** (0.043)
$ln(Income_i)$	−0.0087 (0.013)	−0.0188 (0.015)	0.0003 (0.019)	0.0386* (0.022)	0.0058 (0.014)
Area Fixed Effect	Yes	Yes	Yes	Yes	Yes
No. of observations	96	90	90	90	168
AIC	−398.87	−352.66	−315.75	−287.37	−584.77
Adjusted R-squared	0.101	0.101	−0.018	0.100	0.038

Notes: *, **, and *** indicate statistical significances of 0.1, 0.05, and 0.01, respectively. Standard errors are presented in parentheses. AIC stands for Akaike Information Criterion. The area fixed effect has been added to the model.

extension. The sign of β_3 indicates whether the influence of the new transit stations is positive or negative.

- $\varepsilon_{i,t}$ is a random error term.

To exam the time lag effects of the influence of new transit stations, we performed the DID analysis using the time settings shown in Fig. 9.

The involved months in the DID analysis are shown in Table 4:

In addition to the above DID analysis, since the social media data contains a rich source of information, we further conduct the following analysis to check how new transit stations affect tweet activity and tweet sentiment by two different types of tweet users, people who posted tweets in the station-influenced areas both before and after the opening of new stations (which are referred as “sustaining users” in this paper) versus people posted tweets in the station-influenced areas only before or after the opening of new stations (which are referred as “other users” in this paper). To ensure a relatively fair comparison, we use the 6-month Twitter data before and after the opening of MTR stations to conduct the following analysis.[2] On tweet activity, we test whether new transit stations could expand accessibility on a city level. We focus on the users who posted tweets in the treatment group of each influenced area both before and after the opening of transit stations. On tweet sentiment and content, we investigated whether there is a significant difference in tweet sentiment of people who experienced the change of MTR transit stations and people who visited the station-influenced area only before or after. Moreover, in each influenced area, we also analyzed

[2] It should be noted that, owing to data availability limitations, only 5 months of tweets were collected before the operation of Whampoa and Ho Man Tin MTR stations, while 7 months of tweets were collected before the operation of South Horizons, Lei Tung, Ocean Park, and Wong Chuk Hang MTR stations.

Table 8
Tweet sentiment difference-in-difference analysis results for three MTR station-influenced areas.

Influenced area		Immediate	3-month period	6-month period	12-month period	Whole research period
Whampoa & Ho Man Tin	$T_i \times POST_{i,t}$	0.0272* (0.014)	0.0295* (0.017)	0.0394* (0.020)	0.0386 (0.030)	0.0300 (0.022)
	T_i	−0.0463*** (0.012)	−0.0463*** (0.014)	−0.0463*** (0.017)	−0.0463* (0.025)	−0.0463** (0.020)
	$POST_{i,t}$	−0.0001 (0.010)	0.0054 (0.012)	0.0003 (0.014)	0.0108 (0.021)	0.0064 (0.016)
	Num	34	34	34	34	62
	AIC	−170.83	−156.52	−147.00	−119.54	−246.84
	Adjusted R-squared	0.406	0.325	0.217	0.126	0.132
South Horizons & Lei Tung	$T_i \times POST_{i,t}$	−0.0578** (0.024)	−0.0610*** (0.022)	−0.0276 (0.043)	0.0163 (0.042)	−0.0241 (0.034)
	T_i	0.0364* (0.019)	0.0364** (0.017)	0.0364 (0.034)	0.0364 (0.032)	0.0364 (0.03)
	$POST_{i,t}$	0.0126 (0.017)	0.0132 (0.015)	0.0228 (0.030)	0.0148 (0.030)	0.0136 (0.024)
	Num	38	38	38	34	58
	AIC	−141.99	−148.73	−98.26	−90.43	−167.25
	Adjusted R-squared	0.113	0.165	−0.043	0.089	−0.017
Ocean Park & Wong Chuk Hang	$T_i \times POST_{i,t}$	0.0174 (0.019)	0.0184 (0.022)	−0.0053 (0.021)	0.0059 (0.021)	0.0117 (0.019)
	T_i	0.0057 (0.015)	0.0057 (0.017)	0.0057 (0.017)	0.0057 (0.016)	0.0057 (0.017)
	$POST_{i,t}$	0.0118 (0.014)	0.0107 (0.015)	0.0211 (0.015)	0.0136 (0.015)	0.0127 (0.013)
	Num	38	38	38	38	62
	AIC	−157.76	−150.03	−151.97	−153.33	−250.11
	Adjusted R-squared	0.128	0.096	0.006	0.015	0.071

Notes: *, **, and *** indicate statistical significances of 0.1, 0.05, and 0.01, respectively. Standard errors are presented in parentheses. AIC stands for Akaike Information Criterion.

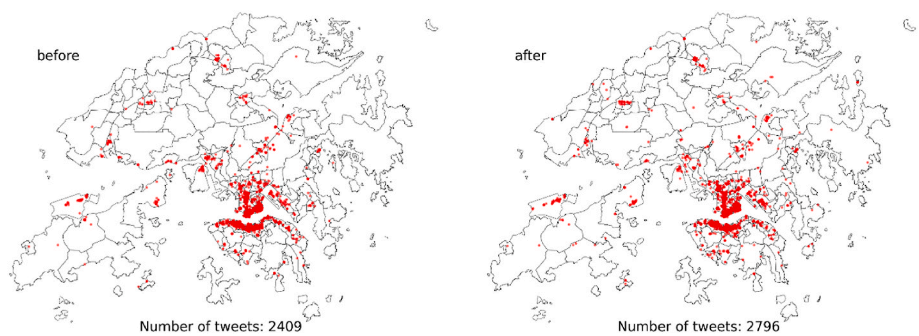

Fig. 12. Footprint comparison of sustaining users in Whampoa & Ho Man Tin.

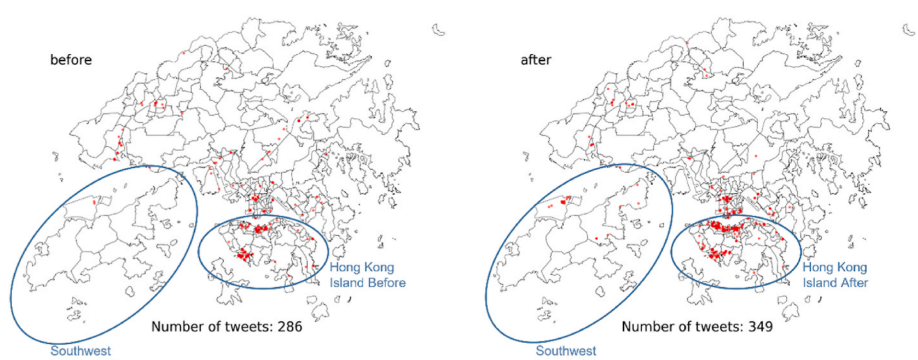

Fig. 13. Footprint comparison of sustaining users - South Horizons & Lei Tung.

the positive tweet text content with respect to two types of users before and after the operation of transit stations.

3.2. Social media data analysis

3.2.1. Tweet sentiment analysis

To determine tweet sentiment, we used a three-class text classification approach, i.e., the sentiment of a tweet could be positive, neutral, or negative. This was performed using *VADER* (Hutto and Gilbert, 2014), a lexicon and rule-based sentiment analysis tool specifically attuned to social media data. Besides estimating the sentiment of words, this tool can also extract sentiment information embedded in special tokens, such as emojis and emoticons, which are significantly popular on social media platforms. For the collected tweets, we firstly used Google Translate's API to translate all the Chinese tweets into English tweets. In the translation process, this API can take both the word meaning and the sequence of words into account. Then, we used *ekphrasis* (Baziotis et al., 2018), a text preprocessing tool designed for social media text, to

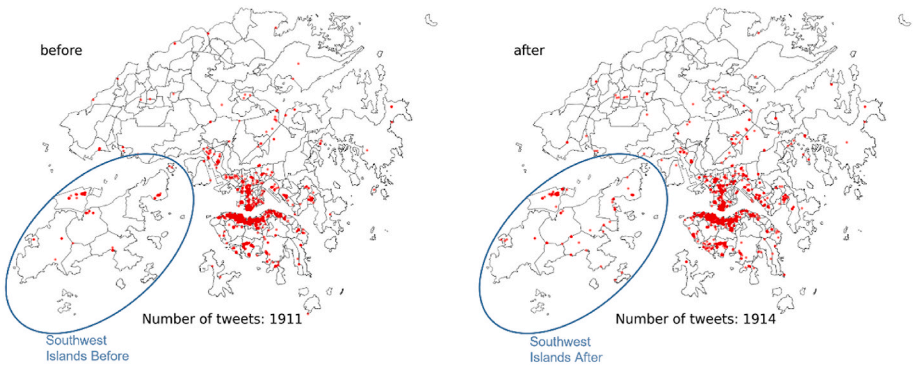

Fig. 14. Footprint comparison of sustaining users - Ocean Park & Wong Chuk Hang.

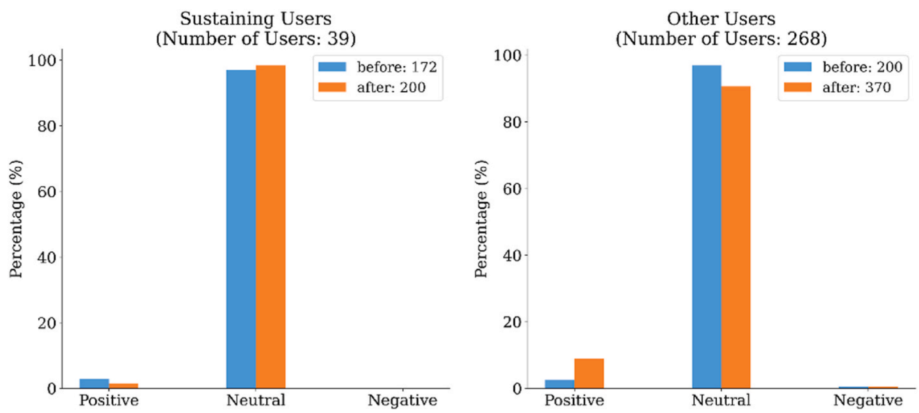

Fig. 15. Sentiment distribution comparison - Whampoa & Ho Man Tin.

Fig. 16. Word-cloud-based positive tweet content comparison - Whampoa & Ho Man Tin.

lowercase all the words and delete and normalize all non-informative tokens, including URLs and numbers. Next, we used *VADER* to analyze the sentiment of each tweet. *VADER* could provide the probability that a tweet is positive, neutral, or negative. To determine the sentiment of each tweet, we chose the sentiment label with the highest probability as the sentiment of the studied tweet.

3.2.2. Tweet text content analysis

We used word clouds to obtain a general view of the words discussed by social media users. A word cloud is a visual representation of text data typically used to depict the word frequency of a large text corpus; the frequency of a word is related to its size in the word cloud.

4. Results and analysis

4.1. Influence of new transit stations on tweet activity

Fig. 10 presents the change in tweet activity level *versus* time in three influenced areas; it should be noted that the tweet activity for each month is recorded as the number of posted tweets per square meter. The horizontal axis represents the time starting from May 2016 to Dec 2018. The black vertical lines indicate the opening month of new MTR line extensions. It should be noted that we remove the last two months of data in South Horizons & Lei Tung since less than ten tweets were found in the treatment group.

From Fig. 10, we observe that before the introduction of MTR stations, the tweet activity of treatment groups of Whampoa & Ho Man Tin, South Horizons & Lei Tung is lower than the control groups. On the

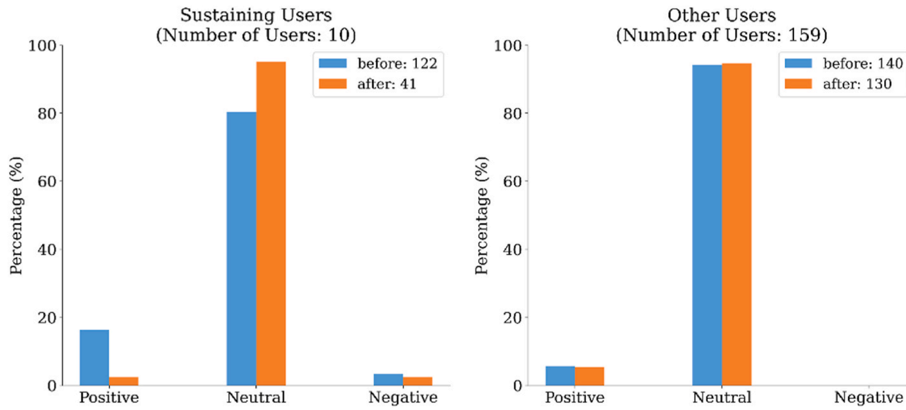

Fig. 17. Sentiment distribution comparison - South Horizons & Lei Tung.

Fig. 18. Word-cloud-based positive tweet content comparison - South Horizons & Lei Tung.

contrary, the tweet activity of the treatment group of Ocean Park & Wong Chuk Hang is higher than the control group. However, the differences between treatment and control groups reduce after the operation of MTR stations. Also, the tweet activity of the treatment group remains relatively stable in the studying period, while that of the control group declines dramatically after the opening of transit stations. These results might indicate that the introduction of MTR stations poses a positive impact on nearby area's tweet activity.

To quantify the relationship between the introduction of new MTR stations and tweet activity, we conducted the DID analysis for three

Fig. 20. Word-cloud-based positive tweet content comparison - Ocean Park & Wong Chuk Hang.

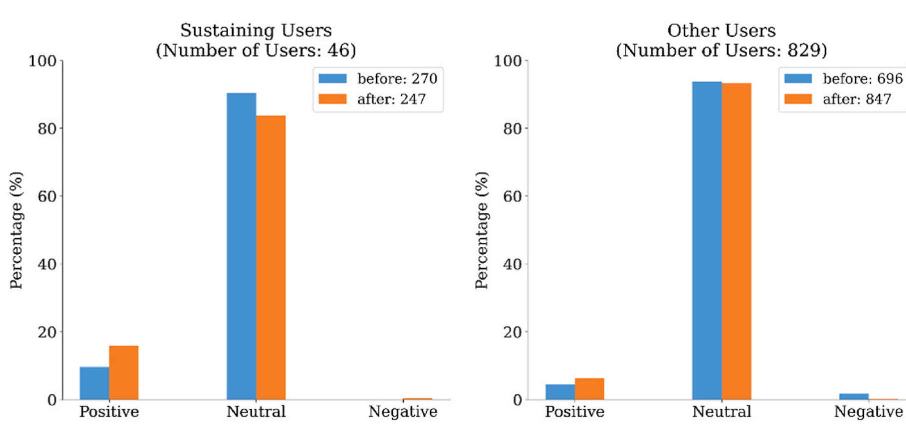

Fig. 19. Sentiment distribution comparison - Ocean Park & Wong Chuk Hang.

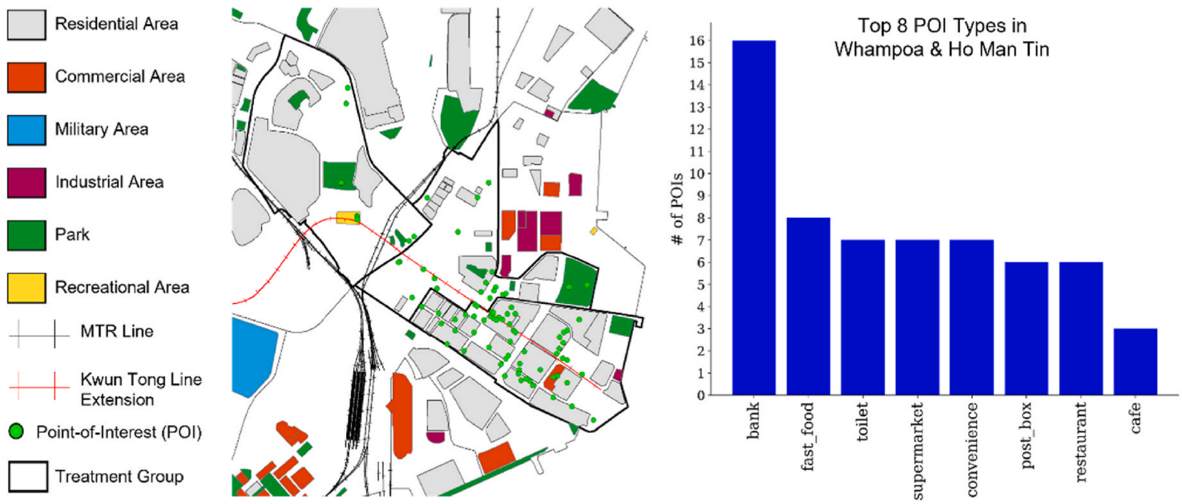

Fig. 21. Land use and POI distribution in Whampoa & Ho Man Tin.

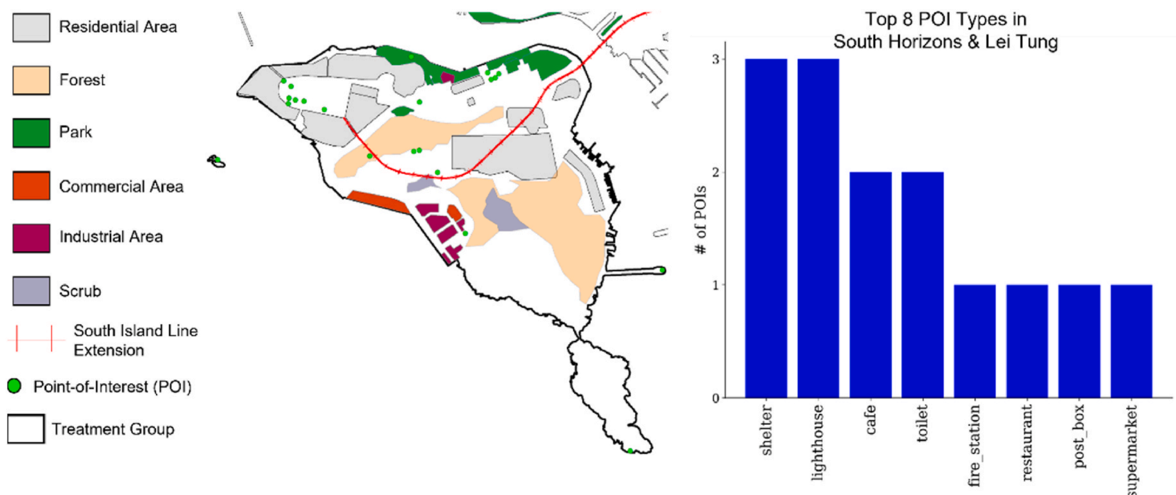

Fig. 22. Land use and POI distribution in South Horizons & Lei Tung.

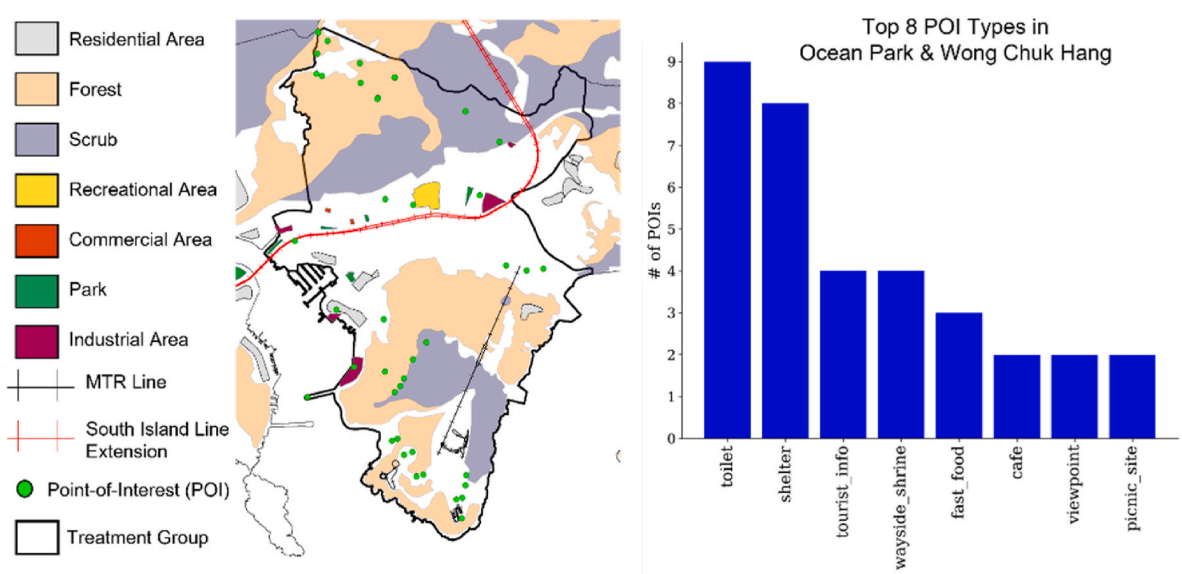

Fig. 23. Land use and POI distribution in Ocean Park & Wong Chuk Hang.

influenced areas. Table 5 shows the result of an overall DID analysis on tweet activity with three influenced areas combined, and Table 6 reports the results of three separate DID analyses, one for each influenced area. The results in Table 5 show that the introduction of new MTR stations has a positive effect on the tweet activity in the nearby area, and the effect is statistically significant after a year of the opening of a new station. The results are controlled for population, income, and area effect. The control variables have negative correlations with tweet activity.

Furthermore, when we separate the analysis into three areas, the results are shown in Table 6. The $T_i \times POST_{i,t}$ variable appears to have a positive coefficient in most time settings, yet no statistical significance has been found. It is also observed that the level of tweet activity decreased gradually for both the treatment group and the control group. This result implies that new transit stations positively influence tweet activity, reverting the downward trend. In these analyses, the control variables, population and income, can no longer be considered because they become constant offset for the control group and the treatment group.

Notes: *, **, and *** indicate statistical significances of 0.1, 0.05, and 0.01, respectively. Standard errors are presented in parentheses. AIC stands for Akaike Information Criterion.

4.2. Influence of new transit stations on tweet sentiment

The changes in tweet sentiment *versus* time in three influenced areas are shown in Fig. 11. The tweet sentiment is measured by the percentage of positive tweets minus the percentage of negative tweets. The horizontal axis represents the time starting from May 2016 to Dec 2018. The black vertical lines indicate the opening month of new MTR line extensions. The horizontal red lines show the sentiment level of all the filtered tweets posted in Hong Kong. These figures show that the introduction of MTR stations has a mixed effect on nearby areas.

The results of the DID analysis for the three areas combined are presented in Table 7. The treatment effect is positive but not significant for the whole study time, suggesting that the introduction of MTR stations has a positive but not significant impact on tweet sentiment of nearby areas. We also found that the tweet sentiment level is higher in areas with a larger population. The difference in income has a limited influence on the tweet sentiment.

To further exam the effect of new MTR stations on tweet sentiment of nearby areas, we conducted the DID analysis between tweet sentiment and the introduction of MTR stations separately for each area. The results are shown in Table 8. In the Whampoa & Ho Man Tin area, the opening of MTR stations has a positive effect on tweet sentiment in the short term, as shown in tests with the time setting of "Immediate", "3-month period", and "6-month period." The treatment group is associated with lower tweet sentiment before the opening of MTR stations, and this trend is reverted right after the opening of MTR stations. However, opposite effects are found in South Horizons & Lei Tung. The tweet sentiment is lowered after the opening of MTR stations in the short term. These short-term effects, positive or negative, become not significant in the long term. In the area of Ocean Park & Wong Chuk Hang, new transit stations pose a positive but no statistically significant impact on the tweet sentiment of station-influenced areas.

4.3. Differences in the effect of new transit stations by user type

We further conduct the following analysis to check how new transit stations affect tweet activity and tweet sentiment by two different types of tweet users, people who posted tweets in the station-influenced areas both before and after the opening of new stations, referred to as sustaining users, versus people posted tweets in the station-influenced areas only before or after the opening of new stations, referred as other users.

First, we test whether new transit stations could expand accessibility on a city level. We focus on the users who posted tweets in the treatment group of each influenced area both before and after the opening of transit stations. We found that the introduction of the MTR transit stations expands the locations that sustaining station-influenced users visit in most cases.

Fig. 12 shows the footprint comparison of these users before and after the introduction of Whampoa and Ho Man Tin MTR stations. We could see that there is not a significant difference in footprints before and after the new MTR stations' opening date.

Fig. 13 shows the footprint comparison of these users in South Horizons & Lei Tung. More footprints were found in other places, such as the southwest islands and Hong Kong Island. This indicates that the introduction of South Horizons and Lei Tung MTR stations may encourage people to visit other places in the city.

In Fig. 14, the footprint comparison of users in Ocean Park & Wong Chuk Hang shows that after the introduction of Ocean Park and Wong Chuk Hang MTR stations, more footprints were found in the southwest islands, people in the nearby area visited more places in the city. This indicates that the introduction of Ocean Park and Wong Chuk Hang MTR stations promotes accessibility.

Moreover, to understand whether there is a significant difference in tweet sentiment of people who experienced the change of MTR transit stations and people who visited the station-influenced area only before or after, we compared the tweet sentiment and content between sustaining station-influenced users and other users before and after the opening of transit stations.

Fig. 15 shows the changes in tweet sentiment for sustaining users and other users in Whampoa & Ho Man Tin. We found that these two new stations attracted new people to visit the station-influenced area, and these people posted more positive tweets.

We also generated word clouds of positive tweets before and after the opening date of Whampoa and Ho Man Tin Stations (Fig. 16). After the introduction of Whampoa & Ho Man Tin MTR stations, besides food and popular place names, words like "festival", "restaurant", and MTR station relevant information appeared in the positive tweets posted by other users, implying that the activities of people who first visit the station influenced area become more diverse.

Fig. 17 depicts the tweet sentiment change of sustaining station-influenced users and other users before and after the operation of South Horizons and Lei Tung MTR stations. There was not a big difference in tweet sentiment of other users, while the tweet sentiment of sustaining station-influenced users was lowered after the opening of new transit stations.

Fig. 18 presents the word clouds of positive tweets posted by two different users before and after the operation of South Horizons and Lei Tung MTR stations. Comparing to the sustaining users, the activities reported by users who first visited South Horizons & Lei Tung become more diverse after the new MTR stations' opening date, since words such as "adventure", "photography", and "store" appear.

In the area of Ocean Park & Wong Chuk Hang, the tweet sentiment change by user type is given in Fig. 19. New transit stations positively influenced both types of users.

Fig. 20 presents the word clouds of positive tweets posted by two different users before and after the operation of Ocean Park and Wong Chuk Hang stations. There was not a significant difference in the tweet content change of positive tweets. A wide range of activities was mentioned by two types of users in both the "before and after" time periods.

5. Discussions and implications on new transit station development

The analysis above suggests that the introduction of the new MTR transit stations generally promotes more tweet activity yet has different effects on people's tweet sentiment for the three influenced areas. While the tweet sentiment in Whampoa & Ho Man Tin is positively influenced by the introduction of new MTR transit stations, the opposite is observed for South Horizons and Lei Tung. This may be due to the different local characteristics of these station-influenced areas. Hence, we review the land use and POI data in the treatment group of each study area and check the possible reasons leading to the difference in tweet sentiment change after the introduction of MTR stations.

In the area of Whampoa & Ho Man Tin, the land use distribution and the number of top 8 types of POIs are given in Fig. 21. The treatment group was mostly composed of residential areas with high density. Moreover, there were many attractive POIs for visits, such as fast-food restaurants and supermarkets. The introduction of MTR stations had positive impacts on nearby area's tweet activity, which may invite more people to visit nearby POIs and positively influence sentiment in social media.

Fig. 22 shows the land use and POI distribution for South Horizons & Lei Tung. Unlike Whampoa & Ho Man Tin, except the residential area, there was also a large proportion of forest and scrub, which had served the local people as a great place for exercise such as walking and jogging for years. Moreover, this area also had a limited number of POIs to serve the local people. However, after the introduction of South Horizons & Lei Tung MTR stations, more people came to visit this place and might use the services provided by the local neighborhood, which may lead to a negative influence on tweet sentiment of the station-influenced area.

The land use and POI distributions of Ocean Park & Wong Chuk Hang are shown in Fig. 23. Unlike the other study areas, this area was mostly composed of forests, scrub, parks, and tourist attractions. Also, Ocean Park Hong Kong, one of the largest theme parks in Asia, is located in this area and directly served by the Ocean Park MTR station. People who visit here would be most likely to take recreational trips. After the introduction of MTR stations, the type of activities in this area did not change greatly. Hence, new MTR stations may pose a positive but not significant impact on nearby area's tweet sentiment.

6. Conclusions

Many studies have reported the positive effects of urban transit stations on public transit share and surrounding property value; however, the influence on nearby area's sentiment in social media remained unclear. This study is the first to evaluate the treatment effect of new transit stations on people's sentiment and activity using Twitter data.

Based on the case of Hong Kong, we focused on three areas influenced by two new MTR line extensions opened in 2016 and estimated whether the introduction of transit stations could lead to significant changes in nearby area's tweet sentiment and tweet activity. When combining the results for the three influenced areas, we found that the transit stations had a positive impact on influenced areas' tweet activity, and the positive impact became significant after six months of the opening of a new station. Moreover, based on the footprint analysis, new transit stations encourage people in nearby areas to visit other places, which suggests that new transit stations can strengthen the connection between the station-influenced areas and other parts of the city. By checking the raw Twitter text, even though new transit stations do not significantly change the types of activities of people who live or work nearby, more people used MTR's service as their travel mode after the introduction of MTR stations.

The influence of new transit stations on tweet sentiment varies by location. In a short term after the opening of new transit stations, some areas showed a positive change in tweet sentiment while some areas showed a negatve change. These short-term effects, positive or negative, become not significant in the long term. This result indicates that the development of the urban transit system in Hong Kong is not necessarily welcomed by all people in nearby areas in a short term.

This study provides new insights into understanding the influence of transit system development on local neighborhoods. The models and tools developed in this study can offer policymakers a quick way to check the local influence of a new transportation project and act as a reference tool for future transportation system development.

However, it is important to highlight the main limitation of this study: Twitter users do not constitute a representative sample of residences in Hong Kong. In future work, traditional methods, e.g. surveys and interviews, should be carried out to study the impact of urban transit stations on people who are not active in social media.

Within the background of the Belt and Road Initiative (BRI), many cross-country infrastructures such as airports and railway stations are being built. The framework presented in this study can help policymakers gather opinions locally and evaluate a project's regional impact. The social media analysis method used in this study, including text mining and sentiment analysis, can be easily extended to multiple language analysis for Singapore, Malaysia, as well as other regions in the belt and road plan. The developed tools could contribute to analyzing the influence of cross-country projects on the local neighborhood in the belt and road plan.

Author contributions

Haoliang Chang: Methodology development, data analysis, writing. Weiran Yao: Data collection and analysis, writing - original draft. Weizun Zhao: validation, writing - review & editing. Jianxiang Huang: Conceptualization, writing - review & editing. Lishuai Li: Study design, project supervision, drafting and final approval of the manuscript.

Declaration of competing interest

None.

Acknowledgements

This work was supported in part by the Hong Kong Research Grants Council Theme-Based Research Scheme under Grant T32-101/15-R, and in part by the General Research Fund under Project 11209717.

Appendix. Tweet filtering process

The tweet filtering process used on the collected social media data in Hong Kong is as follows:

1. Only consider tweets in English and Chinese.
2. Remove verified accounts. An account may be verified if it is determined as an account of public interest (Makice, 2009). These accounts are mostly managed by governments, companies, *etc.*, and are not representative of people's sentiment and an area's people activity.
3. Only keep geocoded tweets.
4. Remove bot accounts. Social media bot accounts can automatically create posts such as advertisements and daily news, which are irrelevant to people's sentiment and an area's people activity level. These accounts should be deleted to produce more reliable results.
5. Remove tweets posted by temporary visitors. In this study, we define Twitter users as temporary visitors if the time span between their first tweet and the last tweet in Hong Kong is less or equal to 7 days.

In the above tweet filtering process, detecting bot accounts is significantly difficult. We firstly used *Botometer* (Davis et al., 2016) to find and filter bots. *Botometer*, which has been used in previous studies (Hasnat and Hasan, 2018), provides the bot-likelihood score of a user by analyzing the user's recent activity. A higher score implies that the user is more likely to be a bot. However, as there is no predefined cut-off value to classify a user as a bot, we manually reviewed a randomly selected sample of user profiles and determined whether they were bot/organizational accounts or humans. From our review process, we finally set the threshold of 0.4 and, thereby, only kept users whose bot likelihood score was less or equal to 0.4. However, after using *Botometer* with a cut-off value of 0.4, we noticed the presence of some bot accounts in the collected data. After analyzing the bot accounts in our dataset, we found the geolocation of tweets posted by these users was almost identical. Hence, we chose to remove accounts which had almost identical latitude-longitude pair in their posted tweets. After testing multiple threshold values, we determined that if 60% of the tweets posted by a user presented almost identical latitude-longitude information, the user was a bot and decided to delete this user.

References

Andrews, F.M., Withey, S.B., 2012. Social Indicators of Well-Being: Americans' Perceptions of Life Quality. Springer, New York.

Baker, D.M., Lee, B., 2019. How does light rail transit (LRT) impact gentrification? Evidence from fourteen US urbanized areas. J. Plann. Educ. Res. 39 (1), 35–49. https://doi.org/10.1177/0739456X17713619.

Bardaka, E., Delgado, M.S., Florax, R.J., 2018. Causal identification of transit-induced gentrification and spatial spillover effects: the case of the Denver light rail. J. Transport Geogr. 71, 15–31. https://doi.org/10.1016/j.jtrangeo.2018.06.025.

Baziotis, C., Pelekis, N., Doulkeridis, C., 2018. DataStories at SemEval-2017 Task 4: Deep LSTM with attention for message-level and topic-based sentiment analysis. In: Proceedings of the 11th International Workshop on Semantic Evaluation. SemEval-2017), pp. 747–754. https://doi.org/10.18653/v1/s17- 2126.

Bertrand, K.Z., Bialik, M., Virdee, K., Gros, A., Bar-Yam, Y., 2013. Sentiment in New York City: A High Resolution Spatial and Temporal View arXiv:1308.5010.

Bollen, J., Mao, H., Pepe, A., 2011. Modeling public mood and emotion: twitter sentiment and socio-economic phenomena. In: Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, pp. 450–453.

Borrego, C., Martins, H., Tchepel, O., Salmim, L., Monteiro, A., Miranda, A.I., 2006. How urban structure can affect city sustainability from an air quality perspective. Environ. Model. Software 21 (4), 461–467. https://doi.org/10.1016/j.envsoft.2004.07.009.

Bowes, D.R., Ihlanfeldt, K.R., 2001. Identifying the impacts of rail transit stations on residential property values. J. Urban Econ. 50 (1), 1–25. https://doi.org/10.1006/juec.2001.2214.

Brown, B.B., Werner, C.M., 2008. Before and after a new light rail stop: resident attitudes, travel behavior, and obesity. J. Am. Plann. Assoc. 75 (1), 5–12. https://doi.org/10.1080/01944360802458013.

Campbell, A., Converse, P.E., Rodgers, W.L., 1976. The Quality of American Life: Perceptions, Evaluations, and Satisfactions. Russell Sage Foundation, New York.

Canales, K.L., Nilsson, I., Delmelle, E., 2019. Do light rail transit investments increase employment opportunities? The case of Charlotte, North Carolina. Reg. Sci. Policy Pract. 11 (1), 189–202. https://doi.org/10.1111/rsp3.12184.

Census and Statistics Department, 2011. 2011 Hong Kong Population Census. https://www.census2011.gov.hk/en/index.html. (Accessed 28 August 2019).

Census and Statistics Department, 2016a. Information Technology Usage Penetration Report 2016. https://www.censtatd.gov.hk/hkstat/sub/sp120.jsp?productCode=C0000031. (Accessed 16 September 2019).

Census and Statistics Department, 2016b. Hong Kong 2016 Population By-Census. https://www.bycensus2016.gov.hk/en/. (Accessed 16 September 2019).

Cervero, R., Murakami, J., 2008. Rail+ Property Development: A Model of Sustainable Transit Finance and Urbanism. Center for Future Urban Transport: A Volvo Center of Excellence, UC Berkeley.

Cervero, R., Murakami, J., 2009. Rail and property development in Hong Kong: experiences and extensions. Urban Stud. 46 (10), 2019–2043. https://doi.org/10.1177/0042098009339431.

Cervero, R., Sullivan, C., 2011. Green TODs: marrying transit-oriented development and green urbanism. Int. J. Sustain. Dev. World Ecol. 18 (3), 210–218. https://doi.org/10.1080/13504509.2011.570801.

Cheshire, P., Sheppard, S., 1995. On the price of land and the value of amenities. Economica 62 (246), 247–267. https://doi.org/10.2307/2554906.

OpenStreetMap Contributors, 2017. OpenStreetMap. https://www.openstreetmap.org. (Accessed 2 September 2017).

Credit, K., 2018. Transit-oriented economic development: the impact of light rail on new business starts in the Phoenix, AZ Region, USA. Urban Stud. 55 (13), 2838–2862. https://doi.org/10.1177/0042098017724119.

Csikszentmihalyi, M., Larson, R., 2014. Validity and reliability of the experience-sampling method. In: Csikszentmihalyi, M. (Ed.), Flow and the Foundations of Positive Psychology: the Collected Works of Mihaly Csikszentmihalyi. Springer, Dordrecht.

Davis, C.A., Varol, O., Ferrara, E., Flammini, A., Menczer, F., 2016. BotOrNot: a system to evaluate social bots. In: Proceedings of the 25th International Conference Companion on World Wide Web, pp. 273–274. https://doi.org/10.1145/2872518.2889302.

Debrezion, G., Pels, E., Rietveld, P., 2011. The impact of rail transport on real estate prices: an empirical analysis of the Dutch housing market. Urban Stud. 48 (5), 997–1015. https://doi.org/10.1177/0042098010371395.

Department, Lands, 2016. Digital Maps. http://www.landsd.gov.hk/mapping/en/digital_map/intro.htm. (Accessed 15 September 2016). Accessed.

Diener, E., Oishi, S., Lucas, R.E., 2003. Personality, culture, and subjective well-being: emotional and cognitive evaluations of life. Annu. Rev. Psychol. 54 (1), 403–425. https://doi.org/10.1146/annurev.psych.54.101601.145056.

Dodds, P.S., Harris, K.D., Kloumann, I.M., Bliss, C.A., Danforth, C.M., 2011. Temporal patterns of happiness and information in a global social network: hedonometrics and Twitter. PLoS One 6 (12), e26752. https://doi.org/10.1371/journal.pone.0026752.

Dubé, J., Legros, D., Thériault, M., Rosiers, F.D., 2014. A spatial difference-in-differences estimator to evaluate the effect of change in public mass transit systems on house prices. Transp. Res. B: Methodol. 64 (6), 24–40. https://doi.org/10.1016/j.trb.2014.02.007.

Duncan, M., 2011. The impact of transit-oriented development on housing prices in San Diego, CA. Urban Stud. 48 (1), 101–127. https://doi.org/10.1177/0042098009359958.

Durahim, A.O., Coşkun, M., 2015. #iamhappybecause: gross national happiness through Twitter analysis and big data. Technol. Forecast. Soc. Change 99, 92–105. https://doi.org/10.1016/j.techfore.2015.06.035.

Frank, M.R., Mitchell, L., Dodds, P.S., Danforth, C.M., 2013. Happiness and the patterns of life: a study of geolocated tweets. Sci. Rep. 3, 2625. https://doi.org/10.1038/srep02625.

Freilich, R.H., 1998. The land-use implications of transit-oriented development: controlling the demand side of transportation congestion and Urban Sprawl. Urban Law 30 (3), 547–572.

Fung, I.C.-H., Tse, Z.T.H., Cheung, C.-N., Miu, A.S., Fu, K.-W., 2014. Ebola and the social media. Correspondence 384 (9961), P2207. https://doi.org/10.1016/S0140-6736(14)62418-1.

González, S.R., Loukaitou-Sideris, A., Chapple, K., 2019. Transit neighborhoods, commercial gentrification, and traffic crashes: exploring the linkages in Los Angeles and the Bay Area. J. Transport Geogr. 77, 79–89. https://doi.org/10.1016/j.jtrangeo.2019.04.010.

Graham, C., 2011. The Pursuit of Happiness: an Economy of Well-Being. Brookings Institution Press, Washington DC.

Gruebner, O., Sykora, M., Lowe, S.R., Shankardass, K., Trinquart, L., Jackson, T., Subramanian, S.V., Galea, S., 2016. Mental health surveillance after the terrorist attacks in Paris. Correspondence 387, 2195–2196. https://doi.org/10.1016/S0140-6736(16)30602-X, 10034.

Hamstead, Z.A., Fisher, D., Ilieva, R.T., Wood, S.A., McPhearson, T., Kremer, P., 2018. 'Geolocated social media as a rapid indicator of park visitation and equitable park access. Comput. Environ. Urban Syst. 72, 38–50. https://doi.org/10.1016/j.compenvurbsys.2018.01.007.

Hasnat, M.M., Hasan, S., 2018. Identifying tourists and analyzing spatial patterns of their destinations from location-based social media data. Transp. Res. C: Emerg. Technol. 96, 38–54. https://doi.org/10.1016/j.trc.2018.09.006.

Hawelka, B., Sitko, I., Beinat, E., Sobolevsky, S., Kazakopoulos, P., Ratti, C., 2014. Geo-located Twitter as proxy for global mobility patterns. Cartogr. Geogr. Inf. Sci. 41, 260–271. https://doi.org/10.1080/15230406.2014.890072.

Helliwell, J.F., Putnam, R.D., 2004. The social context of well-being. Philos. Trans. R. Soc. Lond. B Biol. Sci. 359 (1449), 1435–1446. https://doi.org/10.1098/rstb.2004.1522.

Hess, D.B., Almeida, T.M., 2007. Impact of proximity to light rail rapid transit on station-area property values in Buffalo, New York. Urban Stud. 44 (5–6), 1041–1068. https://doi.org/10.1080/00420980701256005.

Huang, J., Zhang, Q., Li, L., Yang, Y., Chiaradia, A., Pryor, M., Webster, C., 2016. Happiness and high-rise living: sentiment analysis of geo-located Twitter data in Hong Kong's housing estates. In: Proceedings of the 52nd ISOCARP Congress, pp. 380–387.

Hutto, C.J., Gilbert, E., 2014. Vader: a parsimonious rule-based model for sentiment analysis of social media text. In: Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media, pp. 216–225.

Jacobs, J., 1961. The Death and Life of Great American Cities. Random House, New York.

Jones, C.E., Ley, D., 2016. Transit-oriented development and gentrification along Metro Vancouver's low-income SkyTrain corridor. Can. Geogr. 60 (1), 9–22. https://doi.org/10.1111/cag.12256.

Kamruzzaman, M., Baker, D., Washington, S., Turrell, G., 2014. Advance transit oriented development typology: case study in Brisbane, Australia. J. Transport Geogr. 34, 54–70. https://doi.org/10.1016/j.jtrangeo.2013.11.002.

Krefis, A., Augustin, M., Schlünzen, K., Oügge, J., Augustin, J., 2018. How does the urban environment affect health and well-being? A systematic review. Urban Sci. 2 (1), 21. https://doi.org/10.3390/urbansci2010021.

Krueger, A.B., Schkade, D.A., 2008. The reliability of subjective well-being measures. J. Publ. Econ. 92 (8–9), 1833–1845. https://doi.org/10.1016/j.jpubeco.2007.12.015.

Lam, W.H.K., Bell, M.G.H., 2002. Advanced Modeling for Transit Operations and Service Planning. Emerald Group Publishing Limited, Bingley.

Lam, W.H.K., Cheung, C.-Y., Lam, C.F., 1999. A study of crowding effects at the Hong Kong light rail transit stations. Transp. Res. A: Policy Pract. 33 (5), 401–415. https://doi.org/10.1016/S0965-8564(98)00050-0.

Lam, K.-C., Chan, P.-K., Chan, T.-C., Au, W.-H., Hui, W.-C., 2009. Annoyance response to mixed transportation noise in Hong Kong. Appl. Acoust. 70 (1), 1–10. https://doi.org/10.1016/j.apacoust.2008.02.005.

Lansley, G., Longley, P.A., 2016. 'The geography of twitter topics in London'. Comput. Environ. Urban Syst. 58, 85–96. https://doi.org/10.1016/j.compenvurbsys.2016.04.002.

Layard, R., Clark, A.E., Cornaglia, F., Powdthavee, N., Vernoit, J., 2014. What predicts a successful life? A life-course model of well-being. Econ. J. 124 (580), 720–738. https://doi.org/10.1111/ecoj.12170.

Lee, S., Yi, C., Hong, S.-P., 2013. Urban structural hierarchy and the relationship between the ridership of the Seoul Metropolitan Subway and the land-use pattern of the station areas. Cities 35, 69–77. https://doi.org/10.1016/j.cities.2013.06.010.

Lloyd, A., Cheshire, J., 2017. Deriving retail center locations and catchments from geotagged Twitter data. Comput. Environ. Urban Syst. 61A, 108–118. https://doi.org/10.1016/j.compenvurbsys.2016.09.006.

Longley, P.A., Adnan, M., Lansley, G., 2015. The geotemporal demographics of Twitter usage. Environ. Plann. 47 (2), 465–484. https://doi.org/10.1068/a130122p.

Loo, B.P.Y., du Verle, F., 2017. Transit-oriented development in future cities: towards a two-level sustainable mobility strategy. Int. J. Urban Sci. 21, 54–67. https://doi.org/10.1080/12265934.2016.1235488.

Loo, B.P.Y., Chen, C., Chan, E.T.H., 2010. Rail-based transit-oriented development: lessons from New York City and Hong Kong. Landsc. Urban Plann. 97 (3), 202–212. https://doi.org/10.1016/j.landurbplan.2010.06.002.

Lu, M., 1999. Do people move when they say they will? Inconsistencies in individual migration behavior. Popul. Environ. 20 (5), 467–488. https://doi.org/10.1023/A:1023365119874.

Makice, K., 2009. Twitter API: up and Running Learn How to Build Applications with the Twitter API, first ed. O'Reilly Media, Inc., Sebastopol, California.

McGillivray, M., Clarke, M., 2006. Understanding Human Well-Being. United Nations University Press, Tokyo.

Mohan, J., Twigg, L., 2007. Sense of place, quality of life and local socioeconomic context: evidence from the Survey of English Housing, 2002/03. Urban Stud. 44 (10), 2029–2045. https://doi.org/10.1080/00420980701471885.

MTR, 2015. 5 Million Stories Sustainability Report 2015. http://www.mtr.com.hk/en/corporate/sustainability/2015rpt/pdf/mtrfull2015.pdf. (Accessed 10 October 2019). Accessed.

Nadai, M.D., Staiano, J., Larcher, R., Sebe, N., Quercia, D., Lepri, B., 2016. The death and life of great Italian cities: a mobile phone data perspective. In: Proceedings of the 25th International Conference on World Wide Web, pp. 413–423. https://doi.org/10.1145/2872427.2883084.

Nagar, R., Yuan, Q., Freifeld, C.C., Santillana, M., Nojima, A., Chunara, R., Brownstein, J. S., 2014. A case study of the New York City 2012-2013 influenza season with daily geocoded Twitter data from temporal and spatiotemporal perspectives. J. Med. Internet Res. 16 (10), e236. https://doi.org/10.2196/jmir.3416.

National Academies of Sciences, 2005. Engineering, and medicine. In: Transit-oriented Development: Developing a Strategy to Measure Success. The National Academies Press, Washington DC. https://doi.org/10.17226/23319.

Nilsson, I., Delmelle, E., 2018. Transit investments and neighborhood change: on the likelihood of change. J. Transport Geogr. 66, 167–179. https://doi.org/10.1016/j.jtrangeo.2017.12.001.

Oswald, A.J., Powdthavee, N., 2008. Does happiness adapt? A longitudinal study of disability with implications for economists and judges. J. Publ. Econ. 92 (5–6), 1061–1077. https://doi.org/10.1016/j.jpubeco.2008.01.002.

Parkes, A., Kearns, A., Atkinson, R., 2002. What makes people dissatisfied with their neighborhoods? Urban Stud. 39 (13), 2413–2438. https://doi.org/10.1080/0042098022000027031.

Permentier, M., Bolt, G., van Ham, M., 2011. Determinants of neighborhood satisfaction and perception of neighborhood reputation. Urban Stud. 48 (5), 977–996. https://doi.org/10.1177/0042098010367860.

Qi, B., Costin, A., Jia, M., 2020. A framework with efficient extraction and analysis of Twitter data for evaluating public opinions on transportation services. Travel Behav. Soc. 21, 10–23. https://doi.org/10.1016/j.tbs.2020.05.005.

Quercia, D., Ellis, J., Capra, L., Crowcroft, J., 2012. Tracking "gross community happiness" from tweets. In: Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work, pp. 965–968. https://doi.org/10.1145/2145204.2145347.

Rayle, L., 2015. 'Investigating the connection between transit-oriented development and displacement: Four hypotheses. Hous. Pol. Debate 25 (3), 531–548. https://doi.org/10.1080/10511482.2014.951674.

Ryan, S., Frank, L., 2015. Pedestrian environments and transit ridership. J. Public Transp. 12 (1), 39–57. https://doi.org/10.5038/2375-0901.12.1.3.

Seo, K., Golub, A., Kuby, M., 2014. Combined impacts of highways and light rail transit on residential property values: a spatial hedonic price model for Phoenix, Arizona. J. Transport Geogr. 41, 53–62. https://doi.org/10.1016/j.jtrangeo.2014.08.003.

Shao, H., Zhang, Y., Li, W., 2017. Extraction and analysis of city's tourism districts based on social media data. Comput. Environ. Urban Syst. 65, 66–78. https://doi.org/10.1016/j.compenvurbsys.2017.04.010.

Soursourian, M., 2010. Equipping communities to achieve equitable transit-oriented development. Commun. Invest. 22 (2), 22–45.

Statista, 2019. Penetration Rate of Leading Social Networks in Hong Kong from 2nd quarter to 3rd quarter 2018. https://www.statista.com/statistics/412500/hk-social-network-penetration/. (Accessed 10 March 2020).

Wang, S.-W., Sun, L.-S., Rong, J., Yang, Z.-F., 2013. Transit Station Congestion Index Research Based on Pedestrian Simulation and Gray Clustering Evaluation. Discrete Dyn. Nat. Soc., pp. 1–8. https://doi.org/10.1155/2013/891048, 2013.

Wikimedia Commons contributors, "File: Hong Kong Railway Route Map en.pdf," Wikimedia Commons, the free media repository, https://commons.wikimedia.org/w/index.php?title=File:Hong_Kong_Railway_Route_Map_en.pdf&oldid=510507029 (Accessed June 20, 2021).

Wu, W., 2014. Does public investment improve homeowners' happiness? New evidence based on micro surveys in Beijing. Urban Stud. 51 (1), 75–92. https://doi.org/10.1177/0042098013484530.

Yu, H., Jiao, J., Houston, E., Peng, Z.R., 2018. Evaluating the relationship between rail transit and industrial agglomeration: an observation from the Dallas-fort worth region, TX. J. Transp. Geogr. 67, 33–52. https://doi.org/10.1016/j.jtrangeo.2018.01.008.

Zhang, M., 2010. Can transit-oriented development reduce peak-hour congestion? Transport. Res. Rec.: J. Transp. Res. Board 2174 (1), 148–155. https://doi.org/10.3141/2174-19.