What can Social Media Data tell us about the Location and Price of Airbnb Rentals?

Yang Wang*[1], David Philip McArthur†[1] and Jinhyun Hong‡[1]

[1] Urban Big Data Centre, University of Glasgow, Glasgow, G12 8RZ, United Kingdom

January 03, 2019

Summary

With its increasing popularity, Airbnb poses new challenges for planners and policy makers who aim to manage the associated negative externalities. In this paper, we consider how structural attributes, neighbourhood factors and social media activity can explain the location and price of Airbnb rentals. Our contribution is using geotagged tweets to capture the diverse interests of users and to reflect the urban vibrancy or the 'buzz' of different neighbourhoods. The consistent significance of our measure in predicting the price and density of listings shows how fusing new forms of data can better inform policies for the sharing economy.

KEYWORDS: Airbnb, Twitter, Sharing Economy, Neighbourhood, Policy Making, Planning

1. Introduction

The 'sharing economy' has experienced rapid growth over recent years. Airbnb is no-doubt a success in offering 'live-like-a-local' experiences and cheaper short-term rentals than hotels (Tussyadiah and Pesonen, 2016). The presence of tourists in residential environments may however generate negative externalities e.g., noise, safety issues, potentially depressing the local rental market (Gant, 2015, Lee, 2016). A better understanding of the price and spatial distribution of Airbnb listings may help local authorities to manage potential social disruptions, regulate illegal hoteling, and reduce local housing pressure.

The previous investigations of the price and spatial distribution of Airbnb listings is divided into two strands. The first considers the listing prices to provide practical guidance to the landlords (or hosts as they are known in Airbnb parlance). Some determinants of a listing's price are property attributes, hosts' experience, review ratings and photos (Gutt and Herrmann, 2015, Ikkala and Lampinen, 2014, Li et al., 2015). Other research examines neighbourhood factors such as distance to the city centre and tourist attractions, proximity to restaurants, crime rates, and social-economic characters, which can provide valuable information for policy makers (Alizadeh et al., 2018, Quattrone et al., 2016, Xu and Pennington-Gray, 2017).

There is still a challenge in identifying the relevant factors and the models may not be relevant to another context. With a clear marketing strategy of targeting millennials, the market shape of Airbnb shall reflect its peculiar demand.

Meanwhile, social media have proved popular with the millennial generation. Geo-referenced social media data may encapsulate the neighbourhood characteristics esteemed by the younger generations and provide general measures comparable across different cities. Can neighbourhood measures derived from Twitter help explain the variation in Airbnb activities? How important such measures compared

* Yang.Wang@glasgow.ac.uk
† David.Mcarthur@glasgow.ac.uk
‡ Jinhyun.Hong@glasgow.ac.uk

to other neighbourhood factors?

To address these questions, we construct three Twitter variables and compare their impact with other neighbourhood factors in predicting the location and price of Airbnb listing in Chicago. The findings provide evidence that an entropy-based Twitter measure of activity which we believe captures the urban vibrancy or the 'buzz', is significantly correlated with Airbnb activities.

2. Data and Methodology

2.1 Airbnb

Our Airbnb data is a web survey of the Airbnb website from 10-May-2017 in the city of Chicago, Illinois (www.insideairbnb.com). We focus on listings for an entire home/apartment and those with at least one customer review and a non-zero price. Airbnb randomizes the latitude/longitude of listings within 150m. We don't consider this impact on neighbourhood membership.

2.2 Neighbourhood Variables

We treat census tracts (2010) as neighbourhoods. Four neighbourhood factors are calculated using data published on Chicago Data Portal (CDP) including the number of bus, train and metro stations, the aggregated number of registered restaurants plus the number of grocery shops, historical crime records, and proximity to the CBD. We also include median household income from the 2011 American Community Survey (ACS).

2.3 Twitter Data and Variables

We crowdsourced Twitter data by querying tweets within the Chicago bounding box, listening to Twitter's real-time streaming API and accessing historical REST Twitter APIs. This paper draws upon the geo-tagged tweets. We collect 694,338 geo-tagged tweets in Chicago between January 2016 and July 2016.

With the available Tweets, we calculate three Twitter variables in neighbourhoods including 'Twitter_tweets' and 'Twitter_users'. The third variable, 'Twitter_Entropy' is given by Formula 1.

$$Entropy_{Track_i} = -\sum_i^{n_{users}} P(user_i) \log_2 P(user_i), \quad \text{where } P(user_i) = \frac{tweets_{user_i}}{totalTweets_{Track_i}} \quad (1)$$

In a given neighbourhood $Track_i$, with the $totalTweets_{Track_i}$ made by n_{users} and $P(user_i) = \frac{tweets_{user_i}}{totalTweets_{Track_i}}$ the proportion of tweets made by a specific user, $\{P(user_1), P(user_2) \ldots., P(user_n)\}$. The higher the entropy implies that the tweets are distributed evenly across many users. The neighbourhood are likely to be visited by a higher diversity of users. To reduce noise in the data, we removed tweets with the hashtag '#job', '#hiring' and '#career' for the entropy value. Maps c-e in Figure 1 illustrate the Twitter variables in maps.

2.4 Method

We perform regression analyses using Python's 'statsmodels' OLS module to fit the Airbnb price model and Negative Binomial Regressions for the listing model on the neighbourhoods level. Map a,b in Figure 1 shows the aggregated median price and listing counts in neighbourhoods.

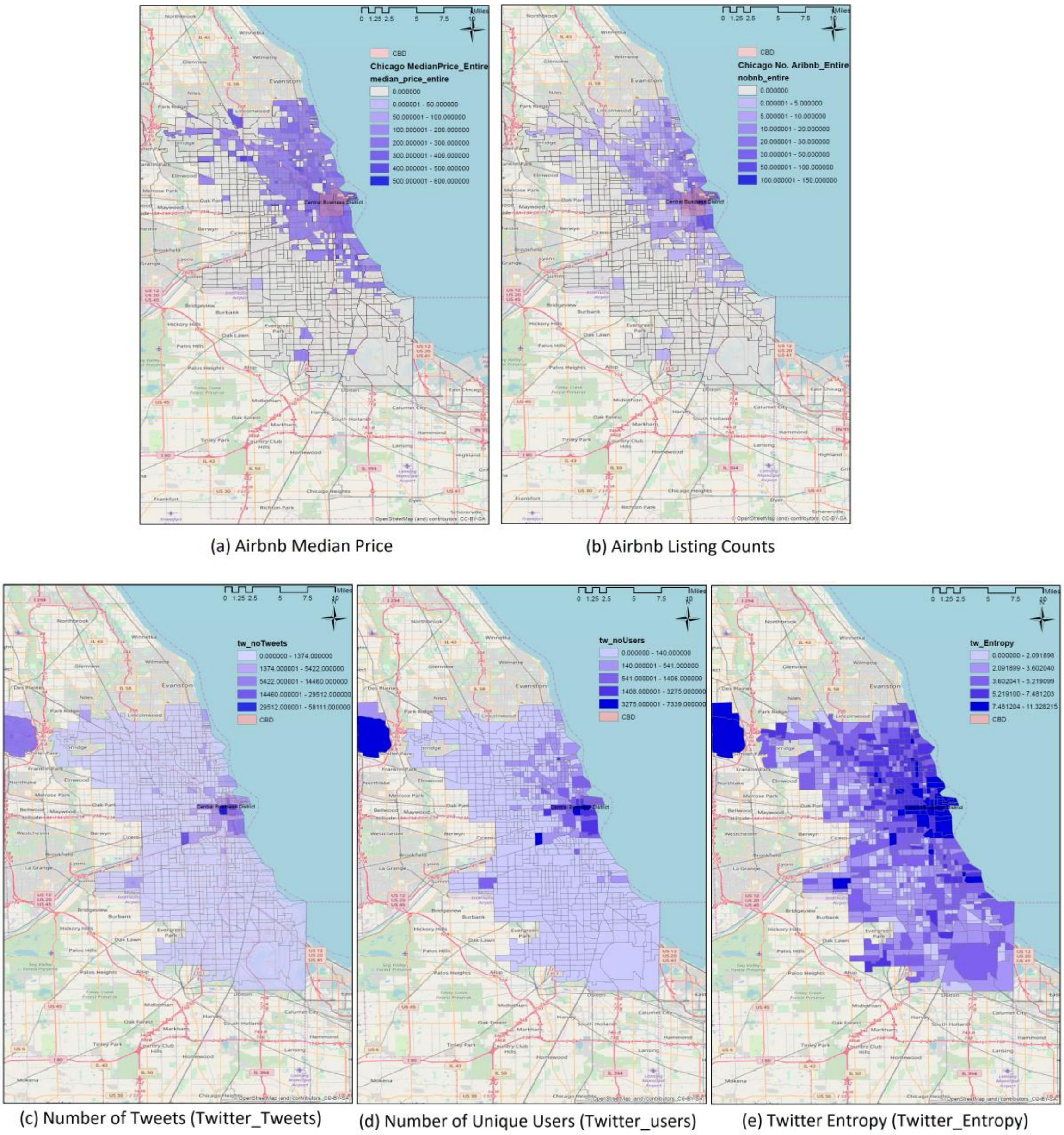

Figure 1. Dependent Variable (a) Airbnb Median Price, (b) Airbnb Listing Counts; Independent Twitter Variables (c) Twitter_tweets (d) Twitter_users, (e) Twitter_Entropy, in neighbourhoods.

3. Result and Discussion

The OLS regression results obtained for the median Airbnb price are illustrated in Table 1. The baseline Model_1 shows that the environmental variables are statistically significant except for the transportation facilities. Introducing the Twitter variables into the model, we find a consistent positive statistical significance throughout Model_2-4. The highest R^2 is seen with the Twitter entropy explaining over 30% the price variance which is very close to the baseline model. Model_5 shows the interesting interactions among the independent variables. The introduction of Twitter entropy reduces the magnitude of the effect of food/grocery places nearly in half. The impact of Twitter entropy confirms to our anticipations that the higher neighbourhood vibrancy the higher the chance that people spend their time and money.

Table 1. OLS and Negative Binomial Regressions results comparing different models for Median Price (Model_1-Model_5) and Listing counts (Model_6-Model_10).

	Model_1	Model_2	Model_3	Model_4	Model_5	Model_6	Model_7	Model_8	Model_9	Model_10
Intercept	3.3183***	-1.0174***	0.2866*	0.2761	1.9988***	1.8044***	-2.3395***	-0.6444***	-0.7123***	0.4047
DistanceToCBD	-0.1875***				-0.1457***	-0.2107***				-0.1602***
	(-12.4045)				(-9.5245)	(-13.2105)				(-10.6849)
MedianHouseholdIncome	0.0197***	0.0198***	0.0338***	0.0347***	0.0125***	0.0109***	0.0119***	0.0246***	0.0267***	0.0051**
	-6.6051	-6.7124	-11.4205	-11.787	-4.1985	-5.0568	-5.4508	-10.4768	-11.0124	-2.4836
	-11.962	(-5.5274)	-1.6924	-1.6242	-6.4928	-7.581	(-14.1278)	(-4.3439)	(-4.7272)	-1.5368
Neighbourhood_busStops	-0.0043				-0.0141	0.0077				-0.0053
	(-0.4245)				(-1.4368)	-0.8872				(-0.6606)
Neighbourhood_crimes	-0.0036***				-0.0032***	-0.0031***				-0.0023***
	(-4.8900)				(-4.5762)	(-4.5688)				(-3.8245)
Neighbourhood_foodGroceryShops	0.0344***				0.0172**	0.0440***				0.0222***
	-4.671				-2.3529	-8.3019				-4.4453
Neighbourhood_metroStation	-0.0224				-0.0611	-0.1354				-0.1779
	(-0.0795)				(-0.2263)	(-0.4780)				(-0.7021)
Neighbourhood_trainStation	0.165				0.0997	-0.0656				-0.1704
	-0.8496				-0.5353	(-0.4619)				(-1.3634)
Twitter_Entropy		0.5329***			0.3862***		0.5348***			0.3553***
		-12.8263			-8.4905		-16.5546			-9.8451
Twitter_tweets				0.0871***					0.1812***	
				-2.7624					-7.7386	
Twitter_users			0.0007***					0.0010***		
			-3.7562					-7.8586		
R2/Psudo R2	0.3575	0.3036	0.1746	0.168	0.4111	0.5238	0.4523	0.2603	0.2317	0.5809
AIC	3426	3481	3617	3623	3358	2787	2825	3146	3024	2670
pearson_chi2						1623	1405	1934	1504	1307
loglikelihood						-1385.984	-1410	-1570.09	-1509.18	-1326.39
N	801	801	801	801	801	801	801	801	801	801
Standard errors in parentheses. * p<.1, ** p<.05, ***p<.01										

The Negative Binomial Regressions depicts the impact of Twitter entropy is similar to the price model but has a stronger relationship with distance to CBD and crime. Twitter entropy has the better-fitted model with AIC closest to baseline. Combining entropy with other neighbourhood factors in Model_10, the coefficients attached to crime, distance to CBD and food/grocery places decrease in magnitude by 24%, 25% and 50%. More listings appear in neighbourhoods which are busy, have low crime, are close to the CBD and which have good accessibility to food places.

4. Remarks and Conclusions

Neighbourhood factors are known to influence the market for Airbnb. Social media data, represented using geo-tagged tweets in this paper, also demonstrates its potential as a significant, general and readily available proxy of neighbourhood characteristics in explaining a significant part of price and distribution of Airbnb listings.

Theoretically, such a measurement may be a good reflection of urban vibrancy (Jacobs, 1961) and the 'buzz' factor (Zukin et al., 2009). Plausibly, the higher the 'buzz', the higher social and economic interactions, the higher attraction of non-residential population, the higher the demand/supply of Airbnb. Practically, the evidence of this paper sheds some light on using social media data to manage the sharing economy by setting policies suitable for particular neighbourhoods.

There are limitations. There is a time gap between our Airbnb and Twitter collection. This reflects the challenges in collecting social media data. With regard to modeling, models at the level of individual listings would allow more potential confounders to be controlled for. Multiple city analyses and seasonal variations are also worth exploring.

References

ALIZADEH, T., FARID, R. & SARKAR, S. 2018. Towards Understanding the Socio-Economic Patterns of Sharing Economy in Australia: An Investigation of Airbnb Listings in Sydney and Melbourne Metropolitan Regions. *Urban Policy and Research*, 1-19.

GANT, A. C. Tourism and commercial gentrification. RC21 International Conference on "The Ideal City: between myth and reality. Representations, policies, contradictions and challenges for tomorrow's urban life" Urbino (Italy), 2015. 27-29.

GUTT, D. & HERRMANN, P. Sharing Means Caring? Hosts' Price Reaction to Rating Visibility. ECIS, 2015.

IKKALA, T. & LAMPINEN, A. Defining the price of hospitality: networked hospitality exchange via Airbnb. Proceedings of the companion publication of the 17th ACM conference on Computer supported cooperative work & social computing, 2014. ACM, 173-176.

JACOBS, J. 1961. *The death and life of American cities*.

LEE, D. 2016. How Airbnb short-term rentals exacerbate Los Angeles's affordable housing crisis: Analysis and policy recommendations. *Harv. L. & Pol'y Rev.*, 10**,** 229.

LI, J., MORENO, A. & ZHANG, D. J. 2015. Agent behavior in the sharing economy: Evidence from Airbnb. *Ross School of Business Working Paper Series*, 1298**,** 2015.

QUATTRONE, G., PROSERPIO, D., QUERCIA, D., CAPRA, L. & MUSOLESI, M. Who benefits from the sharing economy of Airbnb? Proceedings of the 25th international conference on world wide web, 2016. International World Wide Web Conferences Steering Committee, 1385-1394.

TUSSYADIAH, I. P. & PESONEN, J. 2016. Impacts of peer-to-peer accommodation use on travel patterns. *Journal of Travel Research*, 55**,** 1022-1040.

XU, Y.-H. & PENNINGTON-GRAY, L. 2017. Explore the spatial relationship between Airbnb rental and crime.

ZUKIN, S., TRUJILLO, V., FRASE, P., JACKSON, D., RECUBER, T. & WALKER, A. 2009. New retail capital and neighborhood change: boutiques and gentrification in New York City. *City & Community*, 8**,** 47-64.