Spatial patterns of Twitter topics in gentrified areas in Mexico City

Alejandro Sánchez-Zarate
Economic and regional analyst, UN-Habitat Mexico, alsanchez.2332@gmail.com, alsanchez@colmex.mx

1. Introduction

Twitter has risen as a new source of data to study urban phenomena, for example, mobility, segregation, and climate hazards (Cheng et al., 2016, Shelton, 2017, Thakuriah et al., 2017). Several analyses have studied geographical text patterns in global cities such as London and New York. Other branches of study have focused on gentrification processes that are reflected through digital platforms, e.g. Yelp, Foursquare, and AirBnB. However, the active role that digital platforms play to create urban phenomena has not been analyzed. Within this framework, this study seeks to analyze Twitter’s next topic in a gentrified area of Mexico City through a critical digital geographies approach.

2. Data and identification of gentrified zones

First, gentrified neighborhoods were identified by double threshold double threshold—commercial subcentrality method and locations of public-sharing bikes stations.

Next, around 2 million georeferenced tweets in Mexico City were harvested from April 2017 to April 2019 using spritzer stream. The tweets text was pre-processed:

- Deleting stopwords and tweets with 3 words or less
- Identifying named identities linked with places, e.g. “Mexico City”
- Classifying tweets based on where these were emitted (gentrified/non-gentrified zones)
- Classifying tweets by time slots related with consumption practices (evenings and weekends)

3. Topic classification of tweets

1. Unique users’ emission of tweets by time slots and days of the week was identified.
2. Relative Frequency of Word (RFW) by gentrified/non-gentrified zones was calculated.
3. LDA was carried out to classify latent topics from tweets.

4. Descriptive results

Gentrified zones concentrate Twitter users and tweet emissions around peak times for consumption. Emission hotspots were identified outside of gentrified neighborhoods, particularly in shopping centers, airports, and around public transportation. In terms of the main words tweeted by zone, on one hand, gentrified areas showed words related to consumption in iconic, elite parts of the city, e.g. bar, restaurant, happy, café, club, beer, etc. On the other hand, the main words in non-gentrified areas were related to celebrations, concerts, and soccer areas.

5. Twitter topics in gentrified areas

It was carried out LDA algorithm on Twitter text emitted in gentrified neighbourhoods. LDA produced 20 latent topics in gentrified zone. Each latent topic was named based on the main words in each set of words. e.g. hipster or gay culture, creative class, consumption in public space, are positive sentiments. Hotspots of each topic were located in iconic places into gentrified area. It was found out consolidated and transition zones of hipster or bohemian clusters. Also it was recognized co-location of different geography of topics, e.g. overlay hipster, gay and fitness culture on Twitter.

6. Conclusions

This analysis has explored the geography of Twitter topics in Mexico City and its linkage with gentrification zones in the city.

1. Most users and tweets in Mexico City are located in gentrified areas. This implies bias in Twitter as a data source for analyzing urban phenomena.
2. Word patterns from gentry and non-gentry zones were identified.
3. Topics related to gentrification practices, particularly consumption, were found.
4. Future research about the role of digital geography, specifically how code has a role in building augmented realities in gentrified neighborhoods that are consolidated or in-transition.

7. References

