ABSTRACT OF THE THESIS

Developing a Data Mining Framework to Identify a Sense of Gentrification through Social Media Data: A Case Study Using Instagram Posts in Salt Lake City, Utah

by
Cheng-Chia Huang
Master of Science in Geography with a Concentration in Geographic Information Science
San Diego State University, 2017

Gentrification is a transformation of a working-class or abandoned area of a city under the influence of redevelopment and influx of higher-income residents, which involves economic upgrading and replacement of long-term residents who were often of lower social status. Researchers utilize various quantitative and qualitative methodologies to measure the gentrification dynamics; however, incorporating human perceptions of neighborhoods into a large-scale measurement has not been much explored. Moreover, there is a lack of research considering gentrification dynamics at a finer spatio-temporal scale across a large area. This thesis fills these gaps by introducing an innovative social media based data mining framework, which utilizes both qualitative and quantitative approaches by analyzing social media data to explore human activities and perceptions related to gentrification. As a case study, the framework was applied to Instagram data collected from Salt Lake City, UT to investigate gentrification dynamics by capturing the yearly change of a sense of gentrification at a spatial scale of census block groups.

There are two studies in this thesis. The first is a comparative study to examine gentrifying dynamics identified by quantitative and qualitative measures, which are based on five existing census-based gentrification typologies and social media driven *human-perceived gentrifying areas*. As a result, five typologies showed inconsistency in delineating gentrifying areas. Furthermore, they did not match with the gentrifying areas identified based on human perception. The second study introduces a novel data mining framework to examine gentrification dynamics utilizing social media data. Specifically, this study developed two gentrification indicators, *nightlife activities* and *gentrification ambience* to characterize the sense of gentrification. They were extracted from geotagged Instagram posts by employing text processing and text clustering techniques. The spatial distribution of those two indicators over years revealed that areas observed with a sense of gentrification closely correspond to the *human-perceived gentrifying areas*. Furthermore, clustering results delineated two types of gentrifications, residential-driven and commercial-driven gentrifications. Finally, the result portrayed the yearly changes of a sense of gentrification illustrating the potential gentrification expansion. This research demonstrates the framework's capability of incorporating human perception and social media data to explore spatially and temporally fine-grained gentrification dynamics.