Combined GIS Based Spatial-Temporal Analysis Using Social Media Data of Wuhan, China

Uqba Ramzan, Wuhan University, China
Fan Hong, Wuhan University, China

The European Conference on Education 2022
Official Conference Proceedings

Abstract
The development and growth of Internet technology with geo-location has promoted the development of China's Volunteered Geographical Information (VGI) Services. Twitter-like Sina Weibo has gathered a large number of user check-in data, which contains the geolocation features with temporal information. Weibo data has become a major source of geographic location information, helping to access human to service facilities, social events, disaster activities, and real-estate business. This study selects Wuhan (capital of Hubei province) as the study area and combines the collected micro-blog data (2012-2017), POI data (Hubei Surveying and Mapping Bureau) and OSM road network dataset with remote sensing image data. Through spatial inclusion statistical analysis and Change Detection techniques, time and space of Weibo visit frequency and its influence in major universities and commercial pedestrian streets in Wuhan were carried out. This paper will use the clustering algorithm (K-Means), query analysis technique and density analysis method to generate a time-space density cloud of microblog data for institutes and pedestrian streets to find the socio-economical sites within streets and universities, which is beneficial for real-estate business. Both spatial and statistical analysis indicates that the Wuhan University is the university with the highest number of user’s favorite, commentary and content published. All these trend analyses verified through K-mean clustering and change detection techniques to find changes in human mobility patterns of crowds in Wuhan's well-known streets and universities using Google Earth's high-resolution optical imagery. The peak-to-valley analysis of Weibo data reveals past hot events held in Wuhan.

Keywords: VGI, Sina Weibo, Spatio-Temporal, Point of Interest (POI), Open Street Map (OSM), Change Detection, K-Means
INTRODUCTION

With the immense growth and expansion of micro-blog services in China, the growth of Sina Weibo users are increasing with the passage of time. Social media is progressively found in the daily routines of ordinary people in China. The impact of social media is now growing faster and has become an important unseparated part of peoples’ life. With increasing the advancements on the Internet especially on Weibo the mobile phone technology made access to the internet in mobile phones from any location spatially and temporally which can enhance the transmission of online information about different events. Social network analysis uses innovative mathematical methods and statistical analysis to study the associations among users [1]. Social Media (geo-location) data regularly called “check-ins”, about their existence in specific location or venues at given times [2] can be utilized for describing some spatial-temporal consistency in urban areas such as identifying, new buildups and commercial centers [3], detecting local events [4] and determining population distribution [5].

In present, with the expansion of social networks and communication advancements and the popularization of mobile stations, each single user plays the role of a sensor, which in result shows rise in the amount of User-Generated Content (UGC) data (e.g., social media data) to be made obtainable, as well as Volunteer Geographical Information (VGI) [6].

![Figure 1: Data Clock indicating tweet quantity data in one month per hour. (Stefanidis et al. 2012)](image)

A combination of spatial-temporal information gained from social media, public views, outlooks on a hazard extracted from social media and its importance on foreign policy and culture broadcast into an academic area and research can assist government policy-making and help people better recognize the state [7] [8].

**Characteristics and Significance: Social Media Weibo**

Among the Web 2.0 applications, a micro-blog (Weibo in Chinese), a Twitter-like social media has gained substantial popularity in China. This research project will respond to those quantitative questions, collecting the spatial-temporal based social media data from Sina Weibo with containing the textual geographical information, with the help of visual data and statistical analysis to reveal interesting spatial patterns, unusual events, and trends through
various visualization systems [9]. As shown in Fig. 1 the spatial patterns of check-ins using Weibo based on different characteristics like dining, work, home, etc.

**Research Schema and Flow**

The research methodology was designed starting with determining the study area and target virtual community. Since Weibo is the most popular microblog platform in China and this study collected Weibo datasets as a study platform. For this research, essentially Weibo dataset is utilized to understand the content published on social media through spatial-temporal analysis at different times. As the major area is taken of Wuhan which is the capital of the Hubei province of China and the processing of data is shown in Figure 2.

![Figure 2: Structure and Working Flow Chart](image)

The Weibo based analysis was obtained using the different application programming interfaces from 2012 to 2017. After removing redundancy and invalid data, we obtained more than 404454 records in the form of database including the data which published at different places, amounts of content, comment, and likes on different times and dates.

**Overview of study area**

Wuhan (29°58’–31°22’N, 113°41’–115°05’E), the capital of Hubei Province which is most populous city in Central China and one of the nine National Central Cities of China. It is crossed by the Yangtze River which is known as the longest river in China as shown in Figure 3. Wuhan lies on the East of the Jiang-Han plain, a huge area alongside the Yangtze River. Wuhan’s subtropical moist monsoon climate contains four distinct seasons [10]. It is a major transportation hub, with dozens of railways, roads, and expressways passing through the city and connecting to other major cities.
The urban people living density comes to almost 1,200/km² or 3,200 sq mi. Most of the area is plain and decorated with hills and a great number of beautiful universities in which some of the famous universities are taken under study as shown in Figure 4 (a). Besides of lakes and ponds Wuhan is also the hub of transportation due to Yangtze river and contains plenty of shopping streets where most famous commercial streets were chosen for study which are shown in Figure 4 (b).

Figure 4: (a) Famous Universities of Wuhan, Middle: Wuhan University (WHU), Top left: Central China Normal University (CCNU), Top right: China University of Geoscience (CUG), Bottom left: Huazhong Agricultural University (HZAU), Bottom right: Huazhong University of Science and Technology (HUST). (b) Famous Commercial Streets of Wuhan
Considering the physical features of Wuhan, which contains the scenic hills and the beautiful East Lake, Ponds and lush green trees. Wuhan University is recognized as one of the most beautiful and attractive universities in China, particularly for its cherry blossom garden as shown in the center of Figure 4 (a). This campus is occupied by dense and green trees, the view of East Lake and Luojia Hill with sweet-scented flowers everywhere whole year.

Data Acquisition and Processing


In July 2010, Weibo open platform was formally announced and open to the public. Weibo open platform is an information subscription and platform, providing assistance and communication channels with a huge amount of data and Weibo is the most popular microblog platform in China and is dominated by Chinese language users as explained by Louis et al. (2011).

2. Hubei Surveying and Mapping Bureau POI Data

The Point of Interest data downloaded from 湖北测绘局 Hubei Surveying and Mapping Bureau Geolocation data which is utilized with Open Street Map road-network data to analyze maps and graphs concerning time and place.

3. Open Street Map

Open Street Map is a massive database of geographic data, and it’s all open and free. OSM is reflected as a prominent example of volunteered geographic information. On Open Street Map, the content of spatial data is created and controlled by users themselves.

PROPOSED METHODS

In ArcMap workspace, the shapefiles were taken of the following data including the Weibo point data of different times from 2012 to 2017 with attributable properties for example content, like, comment and times of publishing data on social media. The OSM data is also utilized containing corresponding features like the roads, water bodies railways, POI’s (Point of interest), boundary data, land use with different subclasses (forest, residential commercial area, universities, schools, buildings).

1. Buffer Proximity of Streets Boundary: Statistical Analysis

In ArcMap, buffer tool analysis applied in a range which created polygons around input features to a specified distance so that it must cover the area of streets containing commercial markets, shops, and other related POIs. In Figure 5 the selected Green color points indicate the Weibo data taken for all streets and features are clipped using the clip tool. Attribute-based queries are applied for the yearly based like and comment Weibo data analysis.
2. Spatial-Temporal Changes & Trend Mining of Universities & Business Streets with Cloud Density analysis

After taking the boundary around each university and pedestrian streets using buffer proximity and OSM boundary data, the spatial-temporal changes were detected from 2012-2017 through heat map layouts and density algorithms. Figure 6 show the Density Cloud map using Nearest Neighbour and Natural Breaks and stretched from high to low values to heist values and class values density map of the street. Weibo data of each university and street give us the information about the content published which show the most visited, favourite, and commentary sites within the study area.

3. MATLAB K-Means Clustering Algorithm

The clustering of point data of Weibo (2012-2017) is acquired by using MATLAB and applied K-Means algorithm. The point density cloud showing the high to low dense regions are formed by using python on ArcMap.
RESULTS AND DISCUSSION

Statistical Analysis

1.1 Results of Universities

Moreover, besides the text, there are various types to represent the data in the form of hierarchy and graphs. For any social media data like Sina Weibo, the user’s activity could be seen as individual dots or peaks, and the relationship between the two parameters for example like and comment data for 2012-2017 could be seen as in the form of Map layout, bar chart, and table. From 2012-2017 the peak of comments and likes received for Wuhan University is greater say for comments the peak is 13244-5990 and for likes is 36-14235.

The peaks in the bar chart in Figure 7 clearly show Likes, Comments on content published on Weibo and from 2012-2017 the maximum peak is shown by Wuhan University comparative to other universities.

![Figure 7: Universities with frequently published on Weibo from 2012-17. The abbreviations are; Wuhan University (WHU), Central China Normal University (CCNU), Huazhong University of Science and Technology (HUST), China University of Geoscience (CUG) and Huazhong Agriculture University (HZAU).](image)

The statistical results suggest that WHU is the most favorite, commentary and content published university within Wuhan depending on its ancient architecture and people visits as shown in Table 1. The short forms are used for all 5 universities which are taken under study which are; Wuhan University (WHU), Central China Normal University (CCNU), Huazhong University of Science and Technology (HUST), China University of Geoscience (CUG) and Huazhong Agriculture University (HZAU).
Table 1: Universities of Wuhan with frequently published Like “L” and Comments “C” Data on Weibo from 2012 to 2017.

<table>
<thead>
<tr>
<th></th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
<th>2016</th>
<th>2017</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) WHU (L,C)</td>
<td>1) WHU (L,C)</td>
<td>1) WHU (L,C)</td>
<td>1) WHU (L,C)</td>
<td>1) WHU (L,C)</td>
<td>1) WHU (L,C)</td>
<td>1) WHU (L,C)</td>
</tr>
<tr>
<td>2) CCNU (L,C)</td>
<td>2) HUST (L,C)</td>
<td>2) HUST (L,C)</td>
<td>2) CCNU (L,C)</td>
<td>2) CCNU (L,C)</td>
<td>2) CCNU (L,C)</td>
<td>2) CCNU (L,C)</td>
</tr>
<tr>
<td>3) CUG (L,C)</td>
<td>3) CCNU (L,C)</td>
<td>3) CCNU (L,C)</td>
<td>3) CUG (L,C)</td>
<td>3) HUST (L,C)</td>
<td>3) HUST (L,C)</td>
<td>3) HZAU (L,C)</td>
</tr>
<tr>
<td>4) HUST (L,C)</td>
<td>4) CUG (L,C)</td>
<td>4) CUG (L,C)</td>
<td>4) HUST (L,C)</td>
<td>4) CUG (L,C)</td>
<td>4) CUG (L,C)</td>
<td>4) CUG (L,C)</td>
</tr>
<tr>
<td>5) HZAU (L,C)</td>
<td>5) HZAU (L,C)</td>
<td>5) HZAU (L,C)</td>
<td>5) HZAU (L,C)</td>
<td>5) HZAU (L,C)</td>
<td>5) HZAU (L,C)</td>
<td>5) HUST (L,C)</td>
</tr>
</tbody>
</table>

1.2 Statistical Results Obtained from ‘Business’ Streets
The results based on Wuhan business streets exhibit the amount of content published on Weibo and to identify the various corners of business streets, which are extensively used with the help of peoples’ footprints. The field survey of each street was done on different dates to understand which street is more frequently used in week or weekend. The statistical results of streets are shown in Figure 8.

Figure 8: Commercial Streets with frequently published on Weibo from 2012-17
Secondly, Optics Valley Street consists of various shopping malls and different architectural buildings. Statistical analysis shows that Han street showed the maximum peak of Weibo data obtained in the year of 2012 specifically with comment data. In Optics valley the maximum peak of Weibo data obtained in the year of 2017 is more with comments. The trend of visits through statistical analysis Table 2 explains the favorite, likes and comments of famous business streets in different years.

**Table 2:** Comparison of Likes and Comments of Business Streets where (L) stands for Likes and (C) stands for Comments

<table>
<thead>
<tr>
<th></th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
<th>2016</th>
<th>2017</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hanjie</td>
<td>(L,C)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Jianghan</td>
<td>Optics Valley</td>
<td>Optics Valley</td>
<td>Hanjie</td>
<td>Optics Valley</td>
<td></td>
</tr>
<tr>
<td>Jianghan</td>
<td>(L,C)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Optics Valley</td>
<td>Jianghan</td>
<td>Optics Valley</td>
<td>Hanjie</td>
<td>Optics Valley</td>
<td></td>
</tr>
<tr>
<td>Hubu Alley</td>
<td>(L)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Optics Valley</td>
<td>Jianghan</td>
<td>Optics Valley</td>
<td>Hanjie</td>
<td>Optics Valley</td>
<td></td>
</tr>
<tr>
<td>Optimics Valley</td>
<td>(L,C)</td>
<td>Hubu Alley</td>
<td>Hubu Alley</td>
<td>Hubu Alley</td>
<td>Optimics Valley</td>
<td></td>
</tr>
<tr>
<td>Alley</td>
<td>(C)</td>
<td>Hanjie</td>
<td>Hanjie</td>
<td>Hanjie</td>
<td>Jianghan</td>
<td>Jianghan</td>
</tr>
<tr>
<td>Jianghan</td>
<td>(L)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Clustering Algorithms for Universities**

*Point Cloud Density Analysis*

The density diagram generated based upon the points clusters and process in the study area will be known as the point pattern [12]. This type of data is especially important for study including social media networks, structure, for example, Girvan and Newman, (2002) have built a community structure by using centrality indices to discover community boundaries [13]. The point cloud density analysis of Weibo data within the boundaries of different universities were applied to check the frequently visits depending on favoritism, beauty and content published.
Figure 9: Maps Layout of Weibo Point Cloud density of Universities showing high (red), intermediate (green) and Low (pink) regions.

Through all years (2012-2017) Wuhan University is full of red regions as compared to other universities in Wuhan which explains its beauty and attractiveness. Moreover, statistical analysis on Likes and Comments is carried out which better explains and verifies the favorites, commentary, and content published of the aforementioned place as shown in Table 1.

**K-means Clustering Algorithm**

K-Means for major Universities is shown in Figure 10. This computed cluster refers to a collection of data points aggregated together because of certain similarities. For this purpose, we taken target number $k$ for each cluster, which refers to the number of centroids need in the dataset of each universe.

Figure 10: Clustering through K- Means showing centricity of different colored clusters within Universities.
The ‘means’ obtained from the cluster in the K-means refers to averaging of the Weibo data within each university boundary that is, by using the centrality of clusters. Whereas the centroid is the imaginary or real location representing in the form of a triangular shape in the center of the clusters. Results obtained from K-mean proven that the majority of centroid point find for Wuhan university which predicting the major cluster gathering of Weibo points Data.

*K-means algorithm for Major Wuhan Commercial Streets is shown in Figure 11.* The K-means algorithm in data started with a group of randomly selected Weibo points which are used as the beginning points for every cluster and then perform repetitive calculations to optimize the positions of the centroids for each street boundary.

The dense cloud of the point indicated the zone with major K means clustering in street sites. For Jianghan Street the majority K-cluster centricity found in the middle of the street as shown in Figure 11 (a) and for Hubu Alley majority of the mean-cluster lies at right cross-section of the street as expressed in Figure 11 (b). In Han street, the K-means of cluster varying throughout the street as shown in Figure 11 (c) and for Optics Valley major clustering found at the beginning of the street as shown in Figure 11 (d). All the results obtained through K-mean clustering also verify the results obtained through point cloud density.

*Clustering Algorithm for Pedestrian Commercial Streets*  
The density maps are shown in Figure 12 where red color giving us information about the dense regions, green with intermediate visits and purple with least visited sited within commercial pedestrian streets. The number of clusters indicates mostly visited and for real-estate business benefits.

**Figure 11:** K-Means clustering of Jianghan Street (a), Hubu Alley (b), Han Street (c), and Optics Valley (d) with their multi-color cluster centroids.
<table>
<thead>
<tr>
<th>Year</th>
<th>Optics Valley</th>
<th>Han Street</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td><img src="image" alt="Optics Valley Map 2012" /></td>
<td><img src="image" alt="Han Street Map 2012" /></td>
</tr>
<tr>
<td>2013</td>
<td><img src="image" alt="Optics Valley Map 2013" /></td>
<td><img src="image" alt="Han Street Map 2013" /></td>
</tr>
<tr>
<td>2014</td>
<td><img src="image" alt="Optics Valley Map 2014" /></td>
<td><img src="image" alt="Han Street Map 2014" /></td>
</tr>
<tr>
<td>2015</td>
<td><img src="image" alt="Optics Valley Map 2015" /></td>
<td><img src="image" alt="Han Street Map 2015" /></td>
</tr>
</tbody>
</table>
Similarly, the different color in the map gives us various information about the surrounded land-use area. Results of each street based upon the statistical analysis of Weibo data obtained in the GIS domain through this we can identify which street is more recognize with maximum data content published in different times year.

<table>
<thead>
<tr>
<th>Year</th>
<th>Jianghan Lu</th>
<th>Hubu Alley</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td><img src="image" alt="Jianghan Lu Weibo Point Density Analysis, 2012" /></td>
<td><img src="image" alt="Hubu Alley Weibo Point Data Density Analysis, 2012" /></td>
</tr>
<tr>
<td>2013</td>
<td><img src="image" alt="Jianghan Lu Weibo Point Density Analysis, 2013" /></td>
<td><img src="image" alt="Hubu Alley Weibo Point Data Density Analysis, 2013" /></td>
</tr>
</tbody>
</table>
Spatial-temporal maps layout of different streets from 2012 to 2017 are shown in Figure 12 and Figure 13. The red, yellow green, blue, and purple regions indicate the highest, high, intermediate, low and lowest dense regions respectively within the streets. The red region in all the maps at the same place indicates most visited region or shops of that street which is beneficial for real-estate Business.
Change Detection

i. Change in Development in Wuhan University

One of the Major Landmark (Museum) in Wuhan University named as Wanlin Art Museum was completed in end of year 2015 and many people visited this attractive place in 2016 which showed a denser region in that area in 2016 which exhibit the change in development within Universities as shown in Figure. 14

Figure 14: Source: Google Earth pro, Detection of two Images from Google Earth and the corresponding difference in Point data Weibo clusters in 2013 (left) and 2016 (right)
The change is verified through Google Earth's temporal optical imagery. Similarly, one of the buildings along the main road beside Wuhan University which was can be seen started in 2013 through optical high-resolution imagery which was totally constructed in 2016 which showed a big denser region around that area as shown in Figure 15.

ii. Change in Development of Optics Valley Pedestrian Street
Optics Valley Comparison using Kernel Density between 2012 and 2017. Figure 16 shows the optics valley stretched kernel density indicating the change in development in the middle of the street and histogram shows the change in maximum and mean values with the change in standard deviation values indicating the change in the data.
The difference found in Weibo data at different times of year might give us the clue that there could be some activity that took place in that area. The long term study and through Weibo data analysis clarified that the area was under construction in 2012-2014. After 2015 the Boeing 737 airplane as a restaurant in the Optics Valley Pedestrian Street was placed and newly built Cathedral church around the area of Optics valley made the people taking photos, and publishing their geo-location based data content in the form of likes and comments on Weibo. Optics Valley Pedestrian Street became a hotspot in Wuhan, Hubei Province, China and due to built-up of different architectures and multi-country style buildings people liked to visit the other end of the street from 2016-2017 as shown in Figure 16.

**Major Events Indication through ‘Weibo’**

The one of the main findings include the flowing indication about all those past hot events containing peaks based upon comment count and comparison study of events depending upon the number of comments from 2012-2017 (month-wise).
In 2012 the number of comments was very high even the least number of values in June 2012 is approximately equal to the highest values peaks (red box) in 2016 which indicate the hazards. In mid-June, Wuhan recorded the worst levels of air pollution in a decade. Moreover, different Media and news reports suggest and included pictures and content published on social media of Hubei’s capital covered in a greenish-yellow smog [14], and in 2016 there was extreme precipitation which caused the flood in Wuhan [15]. Moreover, the highest peak in 2012 from October to November which shows the major event in China which were Elections in China from October 2012 to March 2013. Afterward, from July to August the rise in comments can be seen because of the London Olympics 2012 which was from Jul 27, 2012 – Aug 12, 2012. Similarly, in 2014 and 2017 there were FIBA championship [16] and Wuhan Marathon respectively which are shown in Figure 17 (green box).

CONCLUSION

For this study Wuhan is selected as the main research area and combines the collected micro-blog data, POI data and OSM road network dataset with remote sensing image data of the research area from 2012 to 2017, using spatial-temporal analysis. Through spatial analysis and statistical analysis of microblog data carried out in the past six years. The results obtained from Open Street Map indicate that Wuhan University is one of the universities considered to be the favorite, commentary and content release. The spatial and statistical results of different
commercial streets and their comparison show that 2012 was the year of Han Street (completed on September 30, 2011), and Optics Valley showed the maximum number of likes and comments posted in 2017. The clustering algorithm (K-Means) and kernel density analysis techniques were used to generate a time-space density cloud of microblog data for higher-educational institutes and commercial pedestrian streets. The main objective of finding the socio-economical sites was achieved through different density analysis. Higher density cloud of people in the map layouts indicate the business sites within the streets and beneficial for real-estate business. Changes in the movement patterns of crowds and development in Wuhan's well-known streets and universities were detected which were verified using Google Earth's high-resolution imagery. The peak-to-valley analysis of Weibo data reveals past hot events. The analysis shows that the content data based on comment counts from 2012 to 2017 reveals many Spatio-temporal events, including public issues such as human behavior, hot topics, and real-estate business sites which will be deeply explored in future by using natural language processing techniques.

ACKNOWLEDGEMENT

I would like to express my deep gratitude to Professor Hong Fan, my research supervisor, for their patient guidance, enthusiastic encouragement and useful critiques of this research work.

Special thanks to State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University and it's faculty for providing the in-depth knowledge about Remote Sensing and GIS and allowing me to study and conduct this research at this prestigious institute. Finally, I wish to thank my parents for their support and encouragement throughout my study.
REFERENCES


Contact email: uqbaramzan1996@outlook.com
               hfan3@whu.edu.cn