CrossMark

Spatiotemporal Analysis of Housing Prices in China: A Big Data Perspective

Shengwen Li[1] · **Xinyue Ye**[2] · **Jay Lee**[2,3] ·
Junfang Gong[1] · **Chenglin Qin**[4]

Abstract Due to the rapid economic growth and urbanization, China's real estate industry has been undergoing a fast-paced development in recent decades. However, the spatial imbalance between the economic growth in urban and that in rural areas and the excessive growth and fluctuations of house prices in both areas had quickly caught public's attention. Not surprisingly, these issues had become a focus of urban and regional economic research. Efficient and accurate prediction of housing prices remains a much needed but disputable topic. Currently, based on the trends and changes in the financial market, population migration and urbanization processes, numerous case studies have been developed to evaluate the mechanism of real estate's price fluctuations. However, few studies were conducted to examine the space-time dynamics of how housing prices fluctuated from a big data perspective. Using data from China's leading online real estate platform {sofang.com}, we investigated the spatiotemporal

✉ Xinyue Ye
xye5@kent.edu

Shengwen Li
swli@cug.edu.cn

Jay Lee
jlee@kent.edu

Junfang Gong
jfgong@cug.edu.cn

Chenglin Qin
qinchlin@vip.sina.com

[1]   School of Information Engineering, China University of Geosciences, Wuhan, China

[2]   Department of Geography, Kent State University, Kent, OH, USA

[3]   College of Environment and Planning, Henan University, Kaifeng, Henan, China and Department of Geography, Kent State University, Kent, OH, USA

[4]   College of Economics, Jinan University, Guangzhou, China

🖉 Springer

trends of the fluctuations of housing prices in the context of big data. This paper uses spatial data analytics and modeling techniques to: first, identify the spatial distribution of housing prices at micro level; second, explore the space-time dynamics of residential properties in the market; and third, detect if there exist geographic disparity in terms of housing prices. Results from our analysis revealed the space-time patterns of the housing prices in a large metropolitan area, demonstrating the utility of big data and means of analyzing big data.

Keywords Housing price · Space-time · Big data · China

Introduction

China has been undergoing rapid urban growth since 1978. The growth of many cities in China has been characterized by the emergence of a polycentric development model of urban land use and its associated housing market. Rapid urbanization in China had stimulated and fueled the housing markets with ever increasing demands for residential units. The instability of the housing prices was argued to be a social focus that influenced and impacted Chinese society in many ways. Jim and Chen (2006) explored the impacts of key environmental elements on the values of residential housing in Guangzhou, including window orientations, green-space views, floor heights, proximity to wooded areas and water bodies, and exposure to traffic noise. Huang and Yin (2014) argued that natural water resources had the most significant positive effects on property values when forming natural recreation clusters or areas. Chen et al. (2011) examined the determinants of housing prices using time-series and cross-sectional data. The dramatic economic development and urbanization had facilitated the growth of housing market. The soaring housing price had been witnessed in many cities (Deng et al. 2012). Barros et al. (2013) examined the time series behavior of housing prices for 69 cities in China to look at potential bursts of artificially inflated housing markets there. The spatial and temporal dynamics of housing prices had become a popular research field drawing much attention from geographers and economists (Xu and Chen 2012; Wu et al. 2014).

Though there are a burgeoning number of papers on big social data from different perspectives, they were rarely used in real world studies. Even adjusted for inflation the economic growth in urban wealth in China had been impressive. That being said, it should be noted that, behind the economic success in China, many cities also faced serious challenges from an imbalanced economic growth and the potential for intensifying social injustice because of the differences in the level of infrastructure development, proximity to markets, natural resource endowments, demographic factors and policy, both inter and intra cities. The data related to these issues could be big and growing in size as well as complex in content. The current literature on housing market analysis in China had been hampered by incomplete evidence because data could only be drawn from conventional statistical yearbooks which only made data available at aggregated scales.

There exist several limitations to the analysis of the spatiotemporal changes in city-level housing prices in China. Different cities exercise different local laws and rules, such as the limitation of the number of units anyone can

purchase and the amount and level of taxes associated with the purchased unit. Furthermore, the economic and political imbalance among cities also contributed to the geographic disparity in housing prices in different areas of China. Research on housing prices at micro-level (intra city) provided an opportunity for avoiding the side effects by political and economic biases. Detecting the space-time dynamics of housing prices using micro level data could also help to understand the spatial structure of urban land use and could shed some light on the mechanism of urban development. It had been widely recognized that location was the primary determinant of housing prices (Kong et al. 2007). But to what extent the variation of housing prices could be explained by their locations had not been empirically examined from a big data perspective until now.

In this paper, we discuss how we applied geospatial analytical methods to examine the big social data of housing prices in Wuhan, China. The data we used in this study covered all postings related to residential housings in China's leading online real estate platform {sofang.com}. We focused on posts in Wuhan over last 10 years. The main focus in this paper is to examine the effect of big data and geography of the selling prices of residential housings in Wuhan. We also discuss how prices vary in different directions in Wuhan at micro level. The statistical analysis and the empirical results reported in this paper have interesting implications for other cities in China, and to a great extent applicable to cities in other developing countries as well.

Over time, cities gradually emerged through the interaction of people or residents (Jiang and Miao 2014). The expansion of housing market mirrored urban growth in China. Analysis of housing markets that was based on big social data often took a bottom-up approach in data collection. Conventional geographic data collected and maintained from the top down by authorities were usually sampled and aggregated. Therefore, they were often small in sizes and sketchy in content. On the other hand, new data harvested from the Internet could be massive and individually based. Due to the high volumes of data collected that way, they are called 'big data.' For example, social media had become a mundane and well-established communication mechanism in the everyday lives of many people. Conventional urban analysis was often from the top down and imposed by authorities or centralized committees. Alternatively, virtual interactions among urban communities through social media were collected in their original forms and content that they tend to reflect a true picture of urban system dynamics. The ubiquitous connectivity to Internet through mobile devices had also transformed our urban and regional environments into hybrid spaces, where social interactions and communication patterns traversed through physical, digital, or a mix of both spaces. Such community-driven social services empowered people to harness the collective intelligence of their communities both locally and globally as they were traversing everyday life and activities.

This paper uses spatial data analysis and modeling techniques to: first, identify the spatial distribution of prices at micro level; second, explore the space-time dynamics of residential properties on the market; and third, detect the trend of housing price variation across space.

Data Acquisition and Analytical Methods Employed

Data Collection and Data Processing

With the proliferation of Internet and cellphone uses, house hunting has evolved from using 'words of mouth' in the past to heavily replying on searches through real estate websites and information disseminated via microblogs. There are several real estate portals in China, such as www.fang.com, house.sina.com.cn, www.xinhuanet.com/house, www.fcmhw.com, www.house365.com, www.fdc.com.cn, and www.anjuke.com. While each of these websites attracted different amounts of use, the fact that the existence of multiple real estate portals attests to the widespread use of such tools.

Among these, SouFun Holdings Ltd., or **Fang.com**, is the leading and biggest Internet portal for real estate in China in terms of the number of page views and visitors to its websites in 2014. This ranking was according to DCCI (http://www.dcci.com.cn), an independent market research institution commissioned. Through the websites, it provides marketing, e-commerce, listing and other value-added services for China's fast-growing real estate and home furnishing and improvement sectors (Newswire 2015). By January 2015, there were more than 82 million active PC and mobile users. Its website and database contained real estate related content covering more than 350 cities in China, which contained almost all real estate built in the last 10 years in all major city. The website recorded the location of a residential property, as well as its sale history and asking price. This information gives us the opportunity for analyzing spatial and temporal trends in housing market with such big data.

Data on a website can be obtained using crawling techniques or through the website's officially released application programming interfaces (API). Usually, API is the first choice for obtaining data from its associated website. But there are normally limits to such access. For example, the number of data or the number of API calling over a specific time duration is often limited. This is to keep its server in good performance by avoiding overloads due to data downloads. Alternatively, the website owner may want to protect its benefits from his competitors. Even with these limitations, web crawling provides us a feasible way for deriving data from websites with only low costs. Fang.com does not provide API to allow data retrieval. But it provides a webpage for browsing all the resident districts in a map. In addition, each district has a series of webpages on fang.com to list the detail information of the real estates, such as the developer, house condition, total area, history of prices, among others. For downloading this information, a web crawling tool was developed to retrieve the data. This web crawling tool first extracted the list of newly-built districts in a city and their locations from the webpage with the map. After that, we extracted the list of historical prices for each house in each district.

- Using Wuhan City in central China as the study area, a total of 4638 new housing units with 7752 listing prices over the last 10 years (i.e., 1/1/2005 ~ 1/31/2015) were acquired and cleaned. While acquiring the data, we also used social media data to assess how active the regions in and around these housing markets were. To that end, Sina Microblog was and still is the largest social media outlet in China. Akin to a hybrid of Twitter and Facebook, it is being used by approximately 30 % (Rapoza 2011) of Internet users in China, with a market penetration similar to Twitter in the

United States. By December 2013, the number of active microblog users reached 129.1 million per month with daily active users estimated to reach approximately 61.4 million. The number of posts in Sina Microblog (a.k.a. Weibo) was estimated to be more than 2.8 billion in December, 2013. Sina Microblog provides developers with APIs to facilitate users in extending the applications in Weibo. Although data can be collected through the search API in Sina Microblog, there exist limitations to the acquisition of details of the data and the amounts of data that can be acquired. As a workaround, we devised an alternative. First, we split the entire study area into small cells. Each cell has the cell size of 0.04° in latitude by 0.04° in longitude. The search API was then deployed repeatedly with 0.4° increment of latitude and longitude from the smallest latitude and latitude of bounding box of Wuhan to the largest latitude and longitude. The Search radius was set to 3.2 km to cover the entire study area. This enabled us to acquire posts around the center of every cell in every hour. In this manner, data acquired by each deployment of the search API did not exceed the limit of the API but collectively the limitations on the geographic details and data volumes were avoided. Finally, duplicated posts were removed based on the identifiers of posts. The non-geotagged posts were also removed from collected posts. As a result of applying the aforementioned procedures, a total of 66,688 geotagged records for 2 days in Wuhan (June 10 and June 11, 2014) were collected.

Analytical Methods

The spatial patterns of changes in housing prices can sometimes be more complex than what a simple wave-like diffusion model can describe. This is because we now have Internet, which geared the way housing prices change from being determined only by local conditions or through a limited network of contacts as before, to what seems to be free of local determinants nowadays. Consequently, studies of spatial diffusion processes of changes in housing prices need to go beyond the traditional approaches that typically fit the trend of housing market expansion into a one-dimensional S-shape diffusion curve.

To better illustrate the underlying cluster patterns of changes in housing prices, the concepts of natural cities and head/tail breaks (Jiang 2013) were applied. The term, *natural cities*, is referred to locations with spatially clustered geographic events, such as the agglomerated patches aggregated from individual locations of social media users (Jiang and Miao 2014). In a spatial pattern of certain geographic events, *head/tail breaks*, are referred to as locations with low frequencies of the events. In the context of head/tail breaks where geographical events only occurred in low frequencies, those locations were highlighted to signify the improbable nature of their occurrences (Jiang 2013). Due to the distribution of price of residential district being skewed with much more cheap housing units than very expensive ones in a district, we adopted the head/tail breaks to emphasize the clusters of residential districts with higher housing prices. It revealed new insights into the structure and dynamics of natural cities (Jiang and Miao 2014).

Nearest neighbor analysis was used to analyze spatial patterns of geographic phenomena (Lee and Wong 2001). It calculates the average distance between each

point and its nearest point, which is compared to the expected value from a complete spatial randomness distribution. The nearest neighbor index, R, can be calculated as:

$$R = \frac{2\sum_{1}^{N} d_i}{\sqrt{NA}}$$

The value of R may range from 0 for a completely clustered point pattern, to 1 for a randomly distributed point pattern, and to about 2.15 for a spatially regular point pattern (Rossbacher 1986). For the purpose of estimating spatial diffusion processes, Lee et al. (2014) used nearest neighbor ratios and then regression curve estimation to model spatial diffusion processes so that main characteristics of diffusion processes could be detected and distinguished.

Geographically weighted regression (GWR, Brunsdon et al. 1998) was used to derive regression coefficients which disclosed the directions and strengths of the relationships between independent variables and dependent variables. Using geographically weighted regression, such relationships did not stay the same over space. Consequently, localized regression coefficients were calculated to show how such relationships vary over space. The local regression coefficients are potentially collinear even if the underlying exogenous variables in the data generating process are uncorrelated (Wheeler and Tiefelsdorf 2005; Lu et al. 2014). Geographically weighted regression was used in a wide variety of studies, such as ecosystem services (Hu et al. 2015), particulate matter (Chu et al. 2015), crime patterns (Cahill and Mulligan 2007), among others. We used GWR only to explore how the association between the dependent variable and the independent variable in the model change spatially. We do not suggest that independent variable included in the GWR in this study represents all possible influencing factors. Rather, we used this analytic method as a tool to reveal geographical disparity of the association between independent and dependent variables.

Spatial Analysis

Spatial Pattern Analysis

We geocoded all the residential properties in the housing market of the study area as Fig. 1a. All the points of interest, or POI, (updated on 29th April, 2015) located in Wuhan city from Open Street map (http://www.openstreetmap.org) were illustrated in Fig. 1b to provide as a reference.

The central part of Fig. 1a and b is the main urban district of Wuhan with Yangtze River going through the middle of it.

Figure 1a and b indicate that a spatial cluster of residential markets did exist and that most of the residential markets were located in the main urban district. Some secondary clusters also could be seen in the figures. Real estate development usually led the expansion and was often ahead of the expansion of the city's administrative boundary. The two maps in Fig. 1 are very similar, reflecting the housing market in Fig. 1a and urban development in Fig. 1b. They were highly and tightly intertwined.
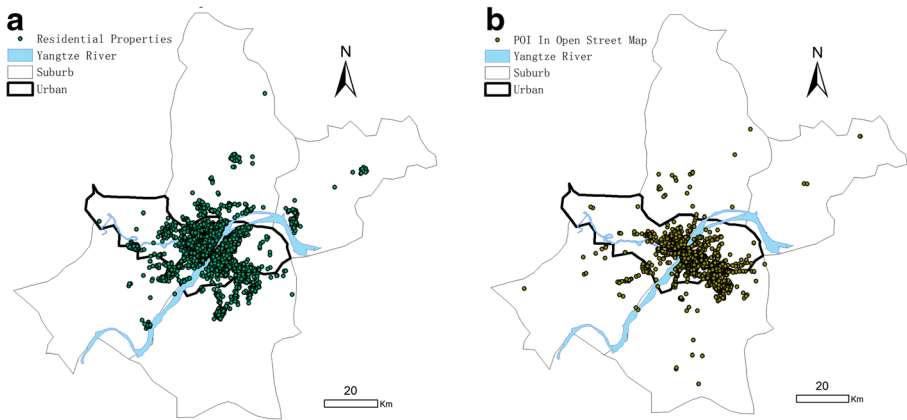
Fig. 1 a The distribution of residential properties in the market over the last 10 years. **b** The distribution f Point of Interest from Open Street Map

Kernel density estimation is a non-parametric method for estimating the probability density function (PDF) of a random point set that had become a widely used analytic method in GIS (Rosenblatt 1956). ArcGIS 10 was used to conduct kernel density estimation for residential properties in the market (Fig. 2a). In the estimation, the search radius was set to 2000 m with a cell size set to 200 m.

Figure 2a indicates the gradient pattern from the city center. In other words, the concentration of housing markets was lessen as the location was increasingly away from the city core. Similar estimations with the same parameter base on the POI from Open Street Map were geocoded and showed in Fig. 2b. Compared to Fig. 2a, b shows that urban functions offered by the city center, though an important factor, was not the only deciding factor that caused the housing prices to vary.
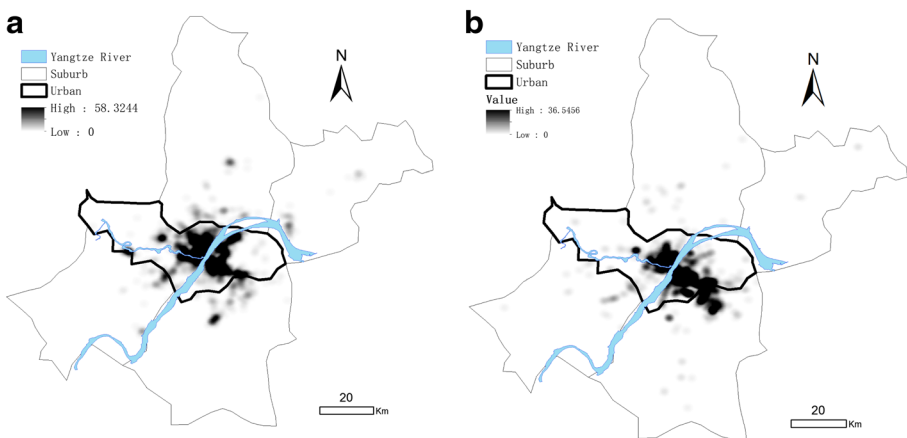

Fig. 2 a Density surface of residential properties in the market over the last 10 years. **b** Density surface of Point-of-Interest from Open Street Map

Spatial Price Dynamics

The housing units in each residential district in Wuhan each had a series of discrete asking prices on different dates. These prices were smoothed by applying moving averages to create a continuous trend of housing prices over time. Given that the sales of housing units did not occur on a regular temporal interval, temporal interpolation was needed to create a consistent temporal profile for each housing unit, or each residential properties. The temporal interpolation was done by setting each month's housing price to be the price of the most recent sale price prior to the month. For example, if a residential property was sold once on 4/25/2008 for $200,000 and again on 5/18/2010 for $250,000, all the monthly housing prices between April 2008 and May 2010 would be set to be $200,000. The monthly housing prices would be set to $250,000 beginning June 2010.

To analyze the spatial patterns of changes in housing prices, we compared the housing prices of 175 residential properties in January 2005, those of 229 in January 2010, and those of 465 in January 2015. We used the Inverse distance weighted interpolation method (IDW) as implemented in ArcGIS to identify how the housing market expanded horizontally (spatially) and how the housing prices grew over time in these areas. The output cell size of IDW processes was set to 200 m and the housing prices were used as the vertical (Z) value field in the IDW spatial interpolation. Figure 3 shows interpolated surfaces at three time periods based on housing prices.

Figure 3 shows the dramatic spatial expansion of housing market. This expansion was partially due to the fast real estate development outside the old urban cores. A major developmental force was due to the fact that the Wuhan East Lake High-Tech Development Zone was accredited as a national independent innovation demonstration zone by the State Council in 2009. The spatial concentration of residential units with high housing prices was apparent throughout the studied period, as indicated by the legend. The skyrocketing housing prices in Chinese housing markets have attracted much global attention (Wu et al. 2012), which was closely related to the much increased urban population and the speedy urbanization in China. Connected to other regions in China by railways and expressways, Wuhan serves as a political, economic, and educational center in Central China. It has witnessed a dramatic economic growth and strong tides of urban construction.
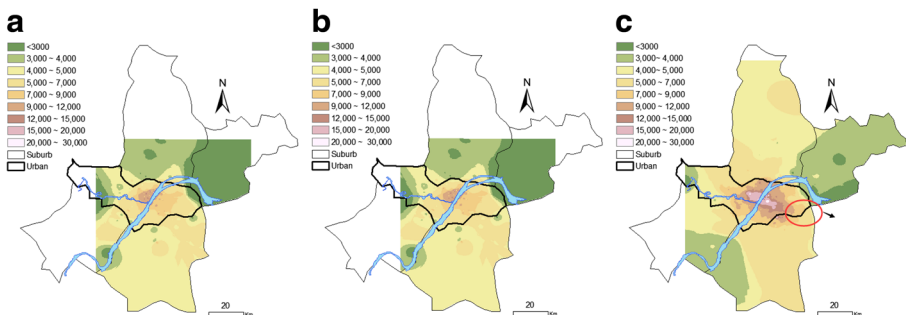

Fig. 3 **a** The price distribution of residential properties in Jan 2005 through IDW. **b** The price distribution of residential properties in Jan 2010 through IDW. **c** The price distribution of residential properties in Jan 2015 through IDW

Head/Tail Breaks Analysis

Head/tail breaks were used to identify the low-frequency data of high housing prices in January 2005, in January 2010, and in January 2015. We classified a residential property as *a head cluster* if its price was higher than the mean housing prices. Similarly, *a tail cluster* was a residential property whose price was lower than the mean housing price of its market. The larger dots (red) in Fig. 4 represent those in head clusters, while the smaller ones indicate ones in the tail clusters. Most of large dots are located inside Wuhan's urban core.

Following the notion of natural cities by Jiang and Miao (2014), it is possible to use a multi-step division of head/tail analysis of housing data to explore the structure and dynamics of a city's residential land use. As shown in Fig. 5, the residential properties clustered beside the Yangtze River in the center of urban area in January 2005. Fast growth spread along Yangtze River from 2005 to 2015. The residential properties in Wuhan clearly experienced a spatial diffusion and witnessed the emergence of multi city centers over time.

Nearest Neighbor Ratio Analysis

If each residential property was represented by a pair of (x,y) coordinates and an associated time stamp (date of sale), it would be possible to calculate a nearest neighbor ratio of a set of such properties at a given time. First, each residential property entered the housing market at different time and stayed in the market until a sale was finalized. Collectively the entries and the exits of all residential properties in the housing market formed a spatial diffusion process. Second, the nearest neighbor ratios calculated from residential properties in the housing market that were on sale at different time periods represented the mechanics of how such spatial diffusion progressed. With this conceptual framework and using Wuhan's data, a series of nearest neighbor ratios were calculated throughout the studies time period. Furthermore, these ratios, forming a trend line over time, were used to fit a non-linear curve to derive a mathematical function that could be used to describe or even predict it.

Using Wuhan data, the nearest neighbor ratios were calculated and then fitted with different regression curves to model the spatial diffusion processes so that main
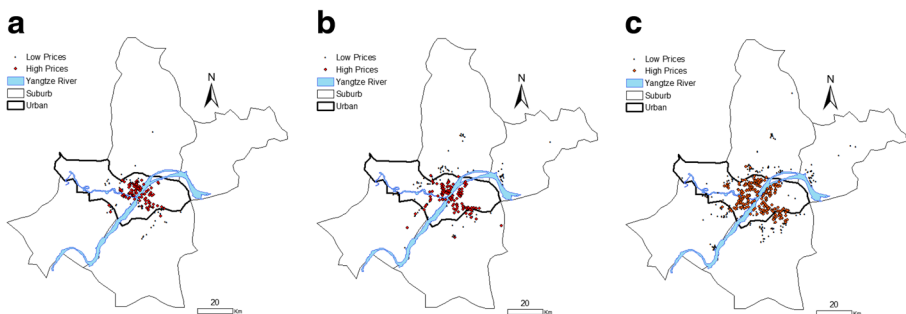

Fig. 4 **a** The price distribution of residential properties in Jan 2005 through Head/tail breaks Analysis. **b** The price distribution of residential properties in Jan 2010 through Head/tail breaks Analysis. **c** The price distribution of residential properties in Jan 2015 through Head/tail breaks Analysis
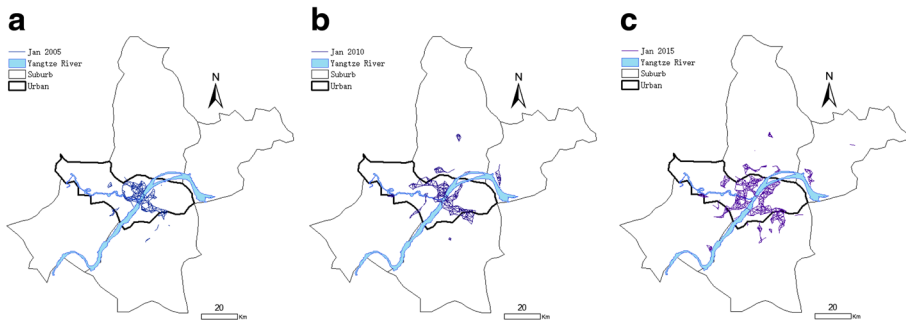
Fig. 5 a Natural cities through housing prices in January 2005. **b** Natural cities through housing prices in January 2010. **c** Natural cities through housing prices in January 2015

characteristics of diffusion processes can be detected and distinguished. Figure 6 shows how the nearest neighbor ratios changed over time. It shows that the degree to which housing prices concentrated spatially had in fact been decreasing over time. That trend suggested that more residential districts had been developed across space (or, more residential land use had seen increases). The estimated curve has some changes in directions, which implies that the development of residential land use might have included some in-fill processes in between built-up areas. When the real estate market soared, the spatial concentration (nearest neighbor ratio) also saw significant increases, such as those in 2008 and 2011.

We used the number of houses (in the housing market) by month to track the spatial diffusion of the housing market. The result is shown in Fig. 6.

Cubic curve is the best with an R-Square of 0.939. According to Lee et al. (2014), relocation diffusion processes are most closely modeled by cubic curves. Consequently, the diffusion of residential house is very likely a relocation diffusion process.

GWR Analysis

According to the above analysis, time and location play important roles in the changes of housing prices. This research collected 65,688 geo-tagged Sina Weibo records (microblogs) on June 10 and 11 (2014). For each residential property in the market, we counted the number of microblogs within the radius of 0.5 mile meters (a possible walking distance for 10 min). Figure 7 shows the relationship between housing price (x) and the number of microblogs (Y). Again, this regression model aims at exploring the relationship between the number of microblogs and the housing prices. It by no means implies that the number of microblogs alone explains the variation in housing prices.

Figure 7 reveals that high housing prices are positively associated to the number of microblogs, which indicates that the areas with high human activity, as characterized by the high volumes of Weibo, tend to be areas of expensive houses. Time-stamped and location-based social media data contributed by individuals constituted a new data source for geographic research. The emerging big data harvested from social media are transforming conventional social sciences into computational social sciences. High human activity via social media also indicates the convenience of accessing urban
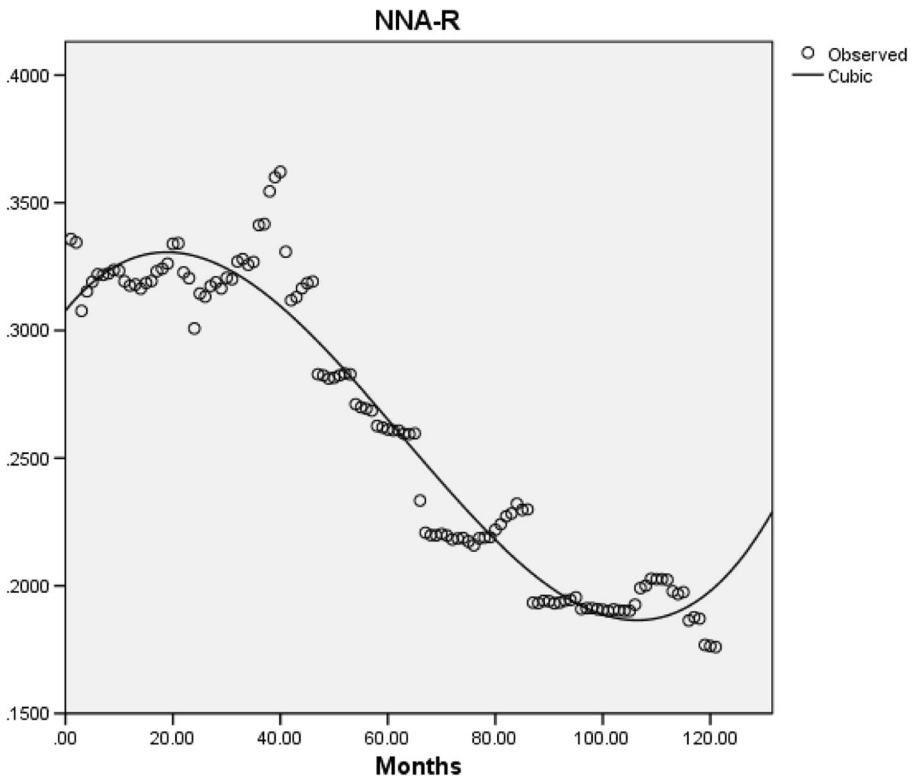
Fig. 6 The spatial diffusion process. Vertical axis shows calculated nearest neighbor ratios. Horizontal axis is a timeline by month

functions and facilities, which is an important indicator of housing prices. In addition, these active areas are mainly areas in CBD, which have very high selling prices or rent.

GWR (Brunsdon et al. 1998) was used to estimate the local effects of the association between the number of microblogs and housing prices. In Fig. 8 the housing price was regressed against the number of microblogs. The city core had a very large residual area shown in red, which indicated the poor fitting of the model in that location. In other words, the customers overpaid the attractiveness of the housing units in these areas from the perspective of the human activity (on social media).
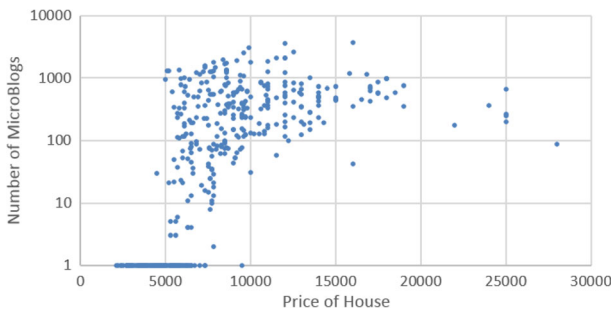

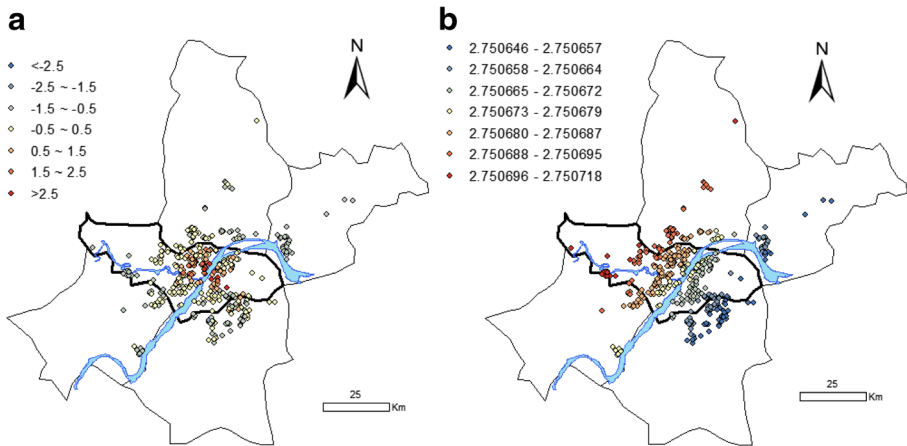**Fig. 7** The price and the number of nearby microblogs

a

- ◆ <-2.5
- ◆ -2.5 ~ -1.5
- ◇ -1.5 ~ -0.5
- ◇ -0.5 ~ 0.5
- ◆ 0.5 ~ 1.5
- ◆ 1.5 ~ 2.5
- ◆ >2.5

N

25 Km

b

- ◆ 2.750646 - 2.750657
- ◆ 2.750658 - 2.750664
- ◇ 2.750665 - 2.750672
- ◇ 2.750673 - 2.750679
- ◆ 2.750680 - 2.750687
- ◆ 2.750688 - 2.750695
- ◆ 2.750696 - 2.750718

N

25 Km

Fig. 8 **a** Standardized residuals of GWR. **b** Regression coefficients for the number of microblogs of GWR

Concluding Remarks

The asking prices of residential houses in Wuhan over the last 10 years were considered in this paper. We found that the dynamics of changes in housing prices were highly relevant to the locations of the housing units. This was especially the case if considering how far the property was away from the city center. In addition, the distribution of residential houses on the market was scattered from the city's perspective, but many local spatial clusters could be identified. Similar housing prices were spatially clustered, driven by the market force. In this paper, we highlighted that emerging social media provided an unprecedented data source for studying the urban activity and consequently allowed a better understanding of the structure and dynamics of real estate markets in cities. Location-based social media, sometimes termed as location-based social networks, referred to a set of Internet-based applications founded on Web 2.0 technologies that was with ideologies of allowing users to create and exchange user-generated content.

Looking into future research, the next-step will be to carry out a comparative space-time analysis of housing prices dynamics among different cities (Ye and Rey 2013). More research is needed for exploring the mechanism of China's housing market and the influencing factors of housing prices (Wang and Kang 2014). Moreover, the quality of school districts as another influencing factor, along with different community attributes and physical environments, should also be considered. Location-based housing data and social media can act as a proxy for socio-economic and environmental conditions of human settlements. They can be used to support a better understanding of the underlying structure and dynamics of the housing markets in the communities. While only microblogs based on two common weekdays and geo-tagged posts were used in the study discussed here, weekend data and posts that were not geo-tagged might also be used in future and further analysis.

References

Barros, C. P., Chen, Z., & Gil-Alana, L. A. (2013). Long memory in the housing price indices in China. *Asian Journal of Empirical Research, 3*(7), 785–807.

Brunsdon, C., Fotheringham, S., & Charlton, M. (1998). Geographically weighted regression. *Journal of the Royal Statistical Society: Series D (The Statistician), 47*(3), 431–443.

Cahill, M., & Mulligan, G. (2007). Using geographically weighted regression to explore local crime patterns. *Social Science Computer Review, 25*(2), 174–193.

Chen, J., Guo, F., & Wu, Y. (2011). One decade of urban housing reform in China: urban housing price dynamics and the role of migration and urbanization, 1995–2005. *Habitat International, 35*(1), 1–8.

Chu, H., Huang, B., & Lin, C. (2015). Modeling the spatio-temporal heterogeneity in the p M10–PM2. 5 relationship. *Atmospheric Environment, 102*, 176–182.

Deng, Y., Gyourko, J. & Wu, J. (2012), *Land and house price measurement in China, Report*, NBER Working Paper No. 18403. National Bureau of Economic Research.

Hu, X., Hong, W., Qiu, R., Hong, T., Chen, C., & Wu, C. (2015). Geographic variations of ecosystem service intensity in Fuzhou City, China. *Science of The Total Environment, 512*, 215–226.

Huang, H., & Yin, L. (2014). Creating sustainable urban built environments: an application of hedonic house price models in Wuhan, China. *Journal of Housing and the Built Environment, 30*(2), 1–17.

Jiang, B. (2013). Head/tail breaks: a new classification scheme for data with a heavy-tailed distribution. *The Professional Geographer, 65*(3), 482–494.

Jiang, B., & Miao, Y. (2014). The evolution of natural cities from the perspective of location-based social media. *The Professional Geographer, 67*(2), 1–12.

Jim, C. Y., & Chen, W. Y. (2006). Impacts of urban environmental elements on residential housing prices in Guangzhou (China). *Landscape and Urban Planning, 78*(4), 422–434.

Kong, F., Yin, H., & Nakagoshi, N. (2007). Using GIS and landscape metrics in the hedonic price modeling of the amenity value of urban green space: a case study in Jinan City, China. *Landscape and Urban Planning, 79*(3–4), 240–252.

Lee, J., & Wong, D. W. S. (2001). *Statistical analysis with ArcView GIS*. New York: Wiley.

Lee, J., Lay, J., Chin, W. C. B., Chi, Y., & Hsueh, Y. (2014). An experiment to model spatial diffusion process with nearest neighbor analysis and regression estimation. *International Journal of Applied Geospatial Research, 5*(1), 1–15.

Lu, B., Charlton, M., Harris, P., & Fotheringham, A. S. (2014). Geographically weighted regression with a non-Euclidean distance metric: a case study using hedonic house price data. *International Journal of Geographical Information Science, 28*(4), 660–681.

Newswire, P. (2015). *SouFun announces five-year plan with a new flagship* Website Fang.com.

Rapoza, K. (2011). China's Weibos vs US's Twitter: And the Winner Is? Forbes. Retrieved 25 August 2015.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics, 27*(3), 832–837.

Rossbacher, L. A. (1986). Nearest-neighbour analysis: a technique for quantitative evaluation of polygonal ground patterns. *Geografiska Annaler. Series A. Physical Geography, 68*(1), 101–105.

Wang, P., & Kang, M. (2014). An empirical analysis on the housing prices in the Pearl River Delta Economic Region of China. *International Journal of Urban Sciences, 18*(1), 103–114.

Wheeler, D., & Tiefelsdorf, M. (2005). Multicollinearity and correlation among local regression coefficients in geographically weighted regression. *Journal of Geographical Systems, 7*(2), 161–187.

Wu, J., Gyourko, J., & Deng, Y. (2012). Evaluating conditions in major Chinese housing markets. *Regional Science and Urban Economics, 42*(3), 531–543.

Wu, J., Deng, Y., & Liu, H. (2014). House price index construction in the nascent housing market: the case of china. *The Journal of Real Estate Finance and Economics, 48*(3), 522–545.

Xu, X. E., & Chen, T. (2012). The effect of monetary policy on real estate price growth in China. *Pacific-Basin Finance Journal, 20*(1), 62–77.

Ye, X., & Rey, S. (2013). A framework for exploratory space-time analysis of economic data. *The Annals of Regional Science, 50*(1), 315–339.