The correlation between Smart-cities and gentrification

Jérémy Scheurer, Philipp Miotti and Rob Spaargaren

***Abstract*— Quick summary of goal and results.**

I. INTRODUCTION

The technological progress of our society shapes our daily life in an unprecedented way and faster than ever before. Since the introduction of the internet, our world has become progressively more connected and new ways of collecting data are introduced continuously. This data is used to implement new systems to improve the quality of life. Following this trend, modern sensors are deployed in cities to communicate with systems (e.g. public transport or parking sensors) to improve the functioning of these systems and offer high quality services to the citizens. A city that deploys many of these systems to improve the general quality of life of its citizens is called a Smart city.

A city which develops itself to a Smart city greatly increases the living standard for its citizens, making it more attractive for people to move there. This spurs an influx of wealthy inhabitants, who also move into the poorer districts of the city, since these districts also experience a great increase in quality of living. These new, wealthy inhabitants invest money to increase the quality of their neighbourhood even further, leading to increasing living costs, driving out the poorer former residents. This process of wealthy inhabitants moving in a neighbourhood, investing money to renovate the neighbourhood, increasing the living quality and living costs and driving out the poorer former residents, is called gentrification. Therefore, there might be a link between gentrification and the development of cities to become Smart. The importance of investigating this link lies in the possibility for cities to pre-emptively change their policies to protect the poorer residents and prevent the neighbourhoods these citizens move into from deteriorating.

To investigate whether this link exists, a quantity indicating the Smartness of a city should be compared to a quantity characterizing the degree of gentrification of the city. To give an indication of the smartness of a city, we will use two publicly available rankings of cities across the globe. The ranking of a city across both lists is used as an indicator of the Smartness of the city. The list we have used are published by the IESE Business School of the Universtiy of Navarra [1], and by the Easypark group [2], an international group specialising in implementation of smart parking in cities.

Initially, gentrification was defined by Ruth Glass in 1964 as the displacement of lower-class working residents in urban neighbourhoods by middle-class citizens [3]. This definition is not useful to quantitatively measure gentrification, so the wide effects of gentrification on a neighbourhood should be utilized. There is a wide variety of effects of gentrification on the neighbourhood, which could be used to quantify it. The nature of the neighbourhood changes, affecting not only housing prices and living costs, but also the look of the neighbourhood, by increasing the amount of public green and parks, and the number of schools. Gentrification also induces a shift in demographics, in culture, educational level and household income. These effects have been used in both qualitative and quantitative research (see Barton, 2016, for an overview [4]). However, the changes in neighbourhood characteristics could be due to a number of other factors like a change of policy by the city government rather than gentrification, and can be subjective of nature. Of the effects on characteristics of the population, an increase of income is the most indicative and least subjective sign of the transition to a higher social class. Using a quantity based on household income to measure gentrification has also been used in other quantitative studies [5]. We present a metric based on the Gross Disposable Household Income (GDHI) that quantifies gentrification in this paper.

II. GENTRIFICATION INDEX AS A METRIC

To quantify gentrification, we have developed a metric based on the GDHI of different districts of a city, which meets certain requirements. Firstly, we require our metric to be independent of the initial state of a city (temporal independence), i.e., to capture merely the gentrification in a certain timespan irrespective of the mutual differences between city parts at an initial time from which the analysis has started. Moreover, it should be possible to compare the extent of gentrification for various cities from all over the world (spatial independence). Thus, we have to make it independent of the local currency and of the overall wealth of a city. In addition, a weighting shall be applied to the neighbourhoods based on their distance. Differences between adjacent and close neighbourhoods will be weighted more heavily than differences between those that are far apart, as the notion of gentrification tries to capture local distinctions. Lastly, we require it to be a non-negative, strictly increasing, and intensive quantity. The last properties imply that the index is zero, if all districts have the same GDHI, and it increases, if the income gap also increases. Intensiveness of our metric means that the it does not scale with the size of the city or with the number of districts.

We propose the change of the GDHI with respect to time as our central quantity to derive our metric. The change of GDHI is compared between the different districts of a city, where the division of a city into districts is by virtue of administrative division units ADU. To study the temportal change of GDHI between different districts across the same timespan, to avoid noise induced by different conditions prevalent to different times, we compared the data of each district in the timespan 2008-2016. How the units for a particular city were actually chosen is described in section III. To assure temporal independence, we calculate the increase of the GDHI in each ADU $x_i = x(District = Districti)$ relative to the increase of the GDHI averaged over the whole city, $\tilde{x}$. More precisely,

$$x_i = \frac{\mathrm{GDHI}(i, 2016) - \mathrm{GDHI}(i, 2008)}{\mathrm{GDHI}(i, 2016)}, \quad \tilde{x} = \frac{\widehat{\mathrm{GDHI}}(2016) - \mathrm{C}}{\widehat{\mathrm{GDHI}}(2\mathrm{C}} \tag{1}$$

where $\widehat{}$ denotes the average of GDHI in a particular year over all districts i. To normalize this increase of income in district i with respect to the increase of income of the rest of the city, and to make it independent of valuta and other city-specific factors, we will use $r_i = (x_i - \tilde{x}) / \tilde{x}$ in our metric. Tab. I shows this relative increase of the GDHI in case of Berlin, which suits as an example to demonstrate our metric. Two normalizations were used in the above definitions: one in x_i or rather $\tilde{x}$, and the other in r_i. Because an absolute gentrification is not sensible, we have to choose relative variables for our gentrification measure. The contribution of the first is twofold; we normalize the GDHI to a specific arbitrary year. This corresponds to a temporal normalization. Without the second "spatial" normalization, we would not be able to compare the cities with each other.

District	GDHI$(i, 2008)$	GDHI$(i, 2016)$	r_i
1	22000	27900	-0.14
2	18000	21300	-0.37
3	17700	25500	$+0.25$
4	19600	24300	-0.21
5	16200	21900	$+0.05$
6	18800	23700	-0.16
7	14800	21600	$+0.27$
8	17400	25800	$+0.32$
9	15500	20400	-0.02
10	17400	21900	-0.17
11	17700	23400	-0.01
12	17800	24600	$+0.13$

TABLE I

MEDIAN HOUSEHOLD INCOME PER MONTH IN EUR IN BERLIN (CITY DIVIDED INTO DISTRICTS ACCORDING TO FIGURE 1) IN YEARS 2008 & 2016, ALONG WITH NORMALIZED INCREASE OF INCOME r_i.

The next step consists of defining a measure for the distance of a district with respect to each other. A generic solution is to use the ADU structure of an city, the districts themselves as distance units. Therefore, we introduce the number of minimal district-border-crossings, which are necessary to get from district i to district j, as our distant measure. For example, the distance, denoted as $\langle \cdot, \cdot \rangle$, of district 1 to district 8 in Fig. 1 is equal to 3, i.e. $\langle 1, 8 \rangle = 3$. Hence, differences in the values r_i of district further apart contribute less to the gentrification index g_i of an particular district i. For the sake of intensiveness we have to normalize this distance by the total number of districts in a city. How

Fig. 1. Map of Berlin's districts with random numbering. Image: © Increa

the distances are calculated exactly is explained in section III.

The final step consists of putting the quantities $\langle \cdot, \cdot \rangle$ and r_i together in a way which results in a meaningful formula for gentrification satisfying our requirements,

$$G = \sum_{i=1}^{N} G_i, \quad G_i = \frac{1}{N} \sum_{j=1,\, j \neq i}^{N} \frac{|r_i - r_j|^2}{\langle i, j \rangle}, \tag{2}$$

We call G and G_i the gentrification index of a city and the gentrification index of the city district i respectively. Due to the fact that gentrification is a rather subjective notion, and the fact that in the perspective of citizens it manifests itself the most when adjacent neighborhoods are compared, we divide by our distance measure $\langle \cdot, \cdot \rangle$. This has the effect to soften the differences between two districts when they are far apart. The meaning of the numerator in Eq. (2) is clear when we write out the squares, $\left(x_i^2 - x_j^2\right) / \tilde{x}^2$. The numerator is smaller, if the mean GDHI of the city $\tilde{x}$ is high. Then, two district with a big difference in the GDHI $x_i - x_j$ are going to make G_i small, whereas on those where the difference is small $\tilde{x}$ has no effect at all. With this feature we can take into account the overall increase of the GDHI in a city. The gentrification index of an city with a low $\tilde{x}$ but a distinctive shift of citizens of different working-classes shall be especially big, as we can then assume that wealthier people haven't moved in from beyond the city at an above average rate.

Where N is the number of districts in the city. We call G and G_i the gentrification index of a city and the gentrification index of district i respectively. Due to the fact that gentrification is a rather subjective notion, and the fact that in the perspective of citizens it manifests itself the most when adjacent neighborhoods are compared, we take the square of our distance measure $\langle \cdot, \cdot \rangle$. This has the effect to soften the differences between two districts when they are far apart. The meaning of the numerator in Eq. (2) is clear when we write out the squares, $\left(x_i^2 - x_j^2\right) / \tilde{x}^2$. The numerator is smaller, if the mean GDHI of the city $\tilde{x}$ is high. Then, two district with a big difference in the GDHI $x_i - x_j$ are going to make G_i small, whereas on those where the difference is small $\tilde{x}$ has no effect at all. With this feature we can take into account the overall increase of the GDHI in a city. The gentrification index of an city with a low $\tilde{x}$ but a distinctive shift of citizens of different working-classes shall be especially big, as we can then assume that wealthier people haven't moved in from beyond the city at an above average rate.

Although, the above argumentation is sensible, we can't be sure if Eq. 2 is the best formula for the index. Thus, we also present some alternatives to G_i, and compare these to the equation. Namely,

$$\begin{aligned} G_i' &= \sum_{j=1,\, j\neq i}^{N} \frac{|r_i - r_j|^2}{\langle i,j\rangle^2} \\ G_i'' &= \sum_{j=1,\, j\neq i}^{N} \frac{|r_i - r_j|}{\langle i,j\rangle^2} \\ G_i''' &= \sum_{j=1,\, j\neq i}^{N} \frac{|r_i - r_j|}{\langle i,j\rangle} \end{aligned} \quad . \tag{3}$$

Tab. II shows the correlation between the gentrification indexes in each district and the quantities r_i for the various types of G_i, and the gentrification index G as well. Based on this results we propose the formula in Eq. (2) as our best measure for the gentrification of a city.

G	0.66	0.51	1.42	1.82
corr $(\cdot, \cdot)$	G_i	G_i'	G_i''	G_i'''
$\lvert r_i \rvert$	0.98	0.96	0.93	0.94

TABLE II

CORRELATION BETWEEN THE DIFFERENT GENTRIFICATION INDICES OF EQUATION **??**.

III. DATA

As described before, our goal consists of comparing a gentrification-index of Smart-cities and Non-Smart-cities. Our first challenge lies in defining which properties make a city Smart or Non-Smart. Secondly we need to calculate the gentrification index which requires the following:

1) Data on the average houshold income per district of a city
2) Data on the distances between districts of a city

In this section we will proceed by explaining our data gathering process for both challenges.

A. Smartness of a city

We explained in our introduction how the term smart city is defined. Also intuitively it is easy to imagine the properties that a smart city can have. Things like a car sharing servics, universal wifi access, a digitalized government, efficient public transport, clean energy generation and much more come to our minds. But to quantitatively asses if a city is smart or not, which means to look at all properties, weigh them accordingly and combine them into a single index, is a very non-trivial task. It would have been out of the scope of our work to attempt to figure out some way of quantizing the Smartness of a city, colllect enough data and create a ranking. We are thus using the smart city rankings of [1] and [2] which use the following criteria to create their rankings:

1 is based on 66 different indicators, subdivided in 10 dimensions: Human capital, social cohesion, economy, public management, governance, mobility and transportation, environment, urban planning, international outreach and techonology.

2 , an international group specialising in implementation of smart parking in cities, have created a ranking based on transport and mobility, sustainability, governance, innovation economy, digitalisation, living standards and expert perception.

The remaining question is how to combine these two rankings when trying to decide between Smart-cities and Non-Smart-cities. When looking at both raknings we can make out, that cities which are ranked in the top Smart cities of one ranking, oftentimes are also ranked as top Smart cities in the other ranking (see table x). Because we lack a ground-truth with which we could compare both rankings, it is very difficult to reason about the differences and about how they come about. Thus we are left only with our intution on which ranking we could trust more. Obviously [1] has several years of experience in producing his "Cities in Motion Index", he also goes into much depth about what properties are relevant for the ranking and why. Finally the university produces an extensive, yearly report about the current rankings. In comparison [2], as a smart parking company, has also much experience in the field and lists all its factors that weigh into the ranking. But on the other hand they go much less into the details of how and why they chose these properties. To sum up, [1] makes definitely a more thorough impression than [2]. But nevertheless [2] is rigorous enough in its analysis, that we decided to keep it aswell as a reference for the Smart-city ranking. Thus for this study we defined a smart city the following way:
A Smart-city, is a city which is ranked in the Top-20 of the Smart-city rankings of [1] and [2].

City	[1] IESE Ranking	[2] Easypark Ranking	Average Ranking
Smart Cities:			
Hamburg	34	14	24
Sydney	16	12	14
Melbourne	14	10	12
New York	1	24	12.5
Berlin	9	13	11
Vienna	15	32	23.5
Chicago	12	36	24
Los Angeles	18	46	32
Boston	4	5	4.5
Washington D.C.	6	28	17
San Francisco	5	7	6
London	2	17	9.5
	Average: 11.3	Average: 20.3	Average: 15.83
Non-Smart Cities:			
San Antonio	N/A	N/A	N/A
San Diego	N/A	N/A	N/A
Austin	N/A	N/A	N/A
Jacksonville	N/A	N/A	N/A
Indianapolis	N/A	N/A	N/A
Columbus	N/A	N/A	N/A
Brisbaine	N/A	N/A	N/A
Adelaide	N/A	N/A	N/A
Perth	N/A	N/A	N/A

For the sake of this paper, we wanted to make the difference between the Smart-cities and the Non-Smart-cities as large as possible. Thus we wanted to define the Non-Smart cities as those which are ranked in the last 20 cities of the rankings of [1] and [2]. Unfortunately we did not find any data on the average houshold income per district of cities like Karachi(Pakistan), Lagos(Nigeria) and others which were at the bottom of the rankings. Because we did find detailed datasets with very high granularity and coverage for countries like the USA and Australia, we choose our Non-Smart cities from these countries. Specifically, we chose cities from these countries which were not listed on the ranking of either [1] or [2] (there were unfortunately no US or Australian cities which were ranked much lower than the Top-20 in either ranking). We argue that this is a reasonable choice for the following reasons. The cities we chose belonged to the largest ones in their corresponding countries, thus one cannot argue that they were purely excluded from the rankings because they were too small and thus not significant. Naturally one could argue that the only reason they did not include these cities is because they did not find any data on them. One could even go further and propose that one of these cities, which we define as Non-Smart, could be in the Top-20 of on e of the the Smart cities rankings. To such an argument we would reason, that by creating such a ranking depending on different properties of cities, by researching what influences the "Smartness" of a city and doing this for several years(this is at least the case for [1]), one could be considered as an expert in the field. And as expert in a field, one could have a good intuition on which city could possibly be a Top Smart-city and which not. Thus it is very unlikely that any of these cities, which are not on the ranking, actually would be a Top-20 city. Because if they were, [1] and [2] would most likely have included them in the ranking. (NOTE I AM INCLUDING THIS WARNING ON PURPOSE IN THE TEXT...THE FOLLOWING ARGUMENT CAN ONLY STAY IN HERE IF OUR RESULTS ACTUALLY SHOW A STRONG DIFFERENCE IN GENTRIFICATION BETWEEN SMART AND NON SMART CITIES). Lastly our results show a strong difference of the gentrification index of Smart-cities and Non-Smart-cities. This makes the case that the cities not ontained in the ranking, actually are Non-Smart Cities even more likely.

B. *Household income and district distance*

For our analysis we need the average household income of a specific district in a city. For some cities the districts are exactly defined by the database(this is mostly the case for European cities) and for other cities we looked at a concise area surrounding its center and defined that as the city(this was mostly the case for US and Australian cities).[1] It is worth mentioning that for all US and Australian cities we used postal codes as "districts" of a city, because they gave us the required granularity that we needed(census tracts were too granular and county subdivisions not granular enough). For European cities(which are usually smaller) having more granular districts was a common trait, thus we just took the districts as defined by the database.

Obviously our household income data comes form very different sources. All of them are official publications form the corresponding cities or countries. Because of the diverse nature of this data, different methods might have been used to estimate the houshold income for a specific district in a city. The data came also in different representations(mean and median), over different time periods(weekly income, monthly income, and yearly income) and for different amount of years(see an overview in Table X). We do realize that we do need to take these statistics with a grain of salt, as they are only estimates. Nevertheless they are the only available data to do any research on. To get the average houshold income into a unified database we did the following preprocessing:

1) If, for a district in a city, we did not have houshold income data for at least half of the years that data was provided, we removed this district from the database altogether.
2) If, for a district in a city, we did not ahve houshold income data for some years(but (1.) did not apply), we manually completed the database in the following way:
    a) If we did have data for a previous and a later year, we took the average of the closest previous and closest later year and set the missing year to that.
    b) If we did have data for a previous year but not for a later year, we set the missing year to the value of the closest previous year.
    c) if we did have data for a later year but not for a previous year, we se the missing year to the value of the closest later year.
3) Lastly extended our dataset to the years of 2008-2016. We need this timeframe to be able to asses the correlation of the Smartness of a city and its gentrification over a longer timeframe, as gentrification is an effect which can only be measured over a longer timeframe. To extend our available data to the mentioned timeframe we fitted a linear polynomial of degree 1 to our data with certain given years. We then just looked at the linear line through our data and calculated its value at certain timestamps we did not currently have data on(see more in Table xx)

As mentioned before, our distance measure for districts is just the shortest path of a certain district to another through other districts. We calculated this manually for all districts in all cities and appended this to our database.[2]

[1] The actual area considered to be in a city can looked up on our gitlab(HERE WEBSITE?) [2] Again the district maps we defined our distances from can be looked up on our gitlab.

City	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	Weekly	Monthly	Yearly	Mean	Median
Smart Cities:																		
Hamburg	✓			✓			✓									✓	✓	
Sydney			✓					✓					✓	✓				✓
Melbourne			✓					✓					✓	✓				✓
New York						✓	✓	✓	✓	✓	✓	✓	✓			✓	✓	
Berlin						✓	✓	✓	✓	✓	✓	✓	✓		✓		✓	
Vienna		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓					✓	✓	
Chicago								✓	✓	✓	✓	✓	✓			✓	✓	
Los Angeles								✓	✓	✓	✓	✓	✓			✓	✓	
Boston								✓	✓	✓	✓	✓	✓			✓	✓	
Washington D.C.								✓	✓	✓	✓	✓	✓			✓	✓	
San Francisco								✓	✓	✓	✓	✓	✓			✓	✓	
London								✓	✓	✓	✓	✓	✓			✓		
Non-Smart Cities:																		
San Antonio								✓	✓	✓	✓	✓	✓			✓	✓	
San Diego								✓	✓	✓	✓	✓	✓			✓	✓	
Austin								✓	✓	✓	✓	✓	✓			✓	✓	
Jacksonville								✓	✓	✓	✓	✓	✓			✓	✓	
Indianapolis								✓	✓	✓	✓	✓	✓			✓	✓	
Columbus								✓	✓	✓	✓	✓	✓			✓	✓	
Brisbaine			✓					✓					✓	✓				✓
Adelaide			✓					✓					✓	✓				✓
Perth			✓					✓					✓	✓				✓

IV. Results

V. CONCLUSIONS

VI. Further Work

For a further improvement of the gentrification index we suggest to incorporate the change of the number of households in each district, and divide the gentrification of that particular district with this change. This also influences gentrification. If, for instance, rental fees increase but the number of household decrease in a certain area, then we can assume that several cheaper households were displaced by a few expensive ones.

Bla

References

[1] Berrone, P., Ricart., J.E. (2017). *IESE Cities in Motion Index*. Retrieved from IESE website: http://www.iese.edu/research/pdfs/ST-0396-E.pdf

[2] Easypark group 2017 Smart Cities Index (n.d.). Retrieved from https://easyparkgroup.com/smart-cities-index/

[3] Ruth Glass (1964). London: aspects of change. London: MacGibbon & Kee.

[4] Barton, M. (2016). An exploration of the importance of the strategy used to identify gentrification, Urban Studies, Vol. 53, Issue 1, pp. 92-111.

[5] Ding, L., Hwang, J., Divringi, E. (2016). *Gentrification and residential mobility in Philadelphia*. Regional Science and Urban Economics, 61, *38-51*.