Spatial Sensitivity Analysis for Urban Hotspots using Cell Phone Traces

Journal Title
XX(X):1–1

DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

Jiahui Wu[1], Enrique Frias-Martinez[2], Vanessa Frias-Martinez[1]

Jiahui Wu is a doctoral candidate in the College of Information Studies at University of Maryland. He received his Master degree in Information Science from Peking University and Bachelor degree in Information Management and Information System from Sun Yat-sen University. His research interests include urban computing, spatio-temporal behavior modeling, machine learning and data visualization.

Enrique Frias-Martinez is a scientific researcher in Telefonica Research, Madrid, Spain. His research areas focus on urban computing, mobility, and big data for social good. Before joining Telefonica, he was a researcher for the Department of Biomedical Engineering of the University of California, Los Angeles, for the Department of Information Systems and Computing of Brunel University (London) and for the Courant Institute of Mathematical Sciences, New York University. Enrique received his PhD in Computer Science from the Universidad Politécnica de Madrid (Spain) in 2002, where he was awarded the Best PhD Thesis Award of the School of Computer Science.

Vanessa Frias-Martinez is an Assistant Professor in the iSchool and UMIACS, and an affiliate Assistant Professor in the Department of Computer Science at the University of Maryland, College Park. Frias-Martinez's research areas are data-driven behavioral modeling and spatio-temporal data mining. Her research focuses on the use of large-scale ubiquitous data to model the interplay between human mobility and the built environment. Specifically, she develops methods to model and predict human behaviors in different contexts and creates tools to aid decision makers in areas such as disaster preparedness and response or urban planning. Her research is funded by the NSF, including a CAREER Award, and by the World Bank. Frias-Martinez received her PhD in Computer Science from Columbia University.

[1]University of Maryland, Maryland, USA
[2]Telefonica Research, Madrid, Spain

Corresponding author:
Jiahui Wu, College of Information Studies, University of Maryland, College Park, MD 20742, USA.
Email: jeffwu@umd.edu

Prepared using sagej.cls *[Version: 2017/01/17 v1.20]*

Spatial Sensitivity Analysis for Urban Hotspots using Cell Phone Traces - Supplementary Materials

Journal Title
XX(X):1–10

DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

Methodology

Spatial units and interpolation methods

Uniform interpolation method: Let $\boldsymbol{v}$ be the set of Voronoi polygons intersected with spatial unit u, v_c be the Voronoi polygon and n_c be the footfall in cell tower c, the interpolated footfall $n_{u,i\text{=Uni}}$ using the Uniform method for u is computed as follows:

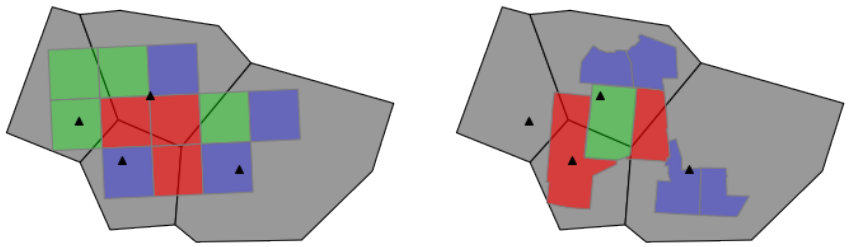
$$n_{u,i\text{=Uni}} = \sum_{v_c \in \boldsymbol{v}} \frac{\text{Area}\left(\text{Intersection}\left(u, v_c\right)\right)}{\text{Area}\left(v_c\right)} \cdot n_c \tag{1}$$

Population-based interpolation method: The population-based method requires information about census population. Let the given census population distribution be at the census tract level, y_x be the shape and p_x be the population of the x-th census tract. The census population is assumed to be normally distributed within y_x, as no finer-grained information about population is available. Let S be an arbitrary polygon *e.g.,* a spatial unit or the intersection between a spatial unit and a Voronoi polygon. If S itself is a census tract, the population of S is straightforward. That is one of the reasons why some researchers prefer to use census tract as spatial units. But if S is not a census tract, then S intersects with a set of census tracts denoted as $\boldsymbol{y}$. The population of S is computed as:

$$\text{Pop}(S) = \sum_{y_x \in \boldsymbol{y}} \frac{\text{Area}\left(\text{Intersection}\left(S, y_x\right)\right)}{\text{Area}\left(y_x\right)} \cdot p_x \tag{2}$$

Let $\boldsymbol{v}$ be the set of Voronoi polygons intersected with spatial unit u, v_c be the Voronoi polygon and n_c be the footfall for cell tower c, the population-based method interpolates the footfall $n_{u,i\text{=Pop}}$ at a given spatial unit u proportional to the population, instead of to the total area:

$$n_{u,i\text{=Pop}} = \sum_{v_c \in \boldsymbol{v}} \frac{\text{Pop}\left(\text{Intersection}\left(u, v_c\right)\right)}{\text{Pop}\left(v_c\right)} \cdot n_c \tag{3}$$



(a) Grids (b) Census tracts

Supp. Fig. 1. An example of grids intersecting with Voronoi polygons (the underlying grey polygons). The locations of cell towers are represented as black triangles. Grids or census tracts in red, green and blue intersect with three, two and one Voronoi polygons, respectively. For example, G_1 intersects with Vor_1 and Vor_2, G_2 intersects with Vor_2, CT_1 intersects with Vor_1 and Vor_3 and CT_2 intersects with Vor_1.

Inverse distance weighting interpolation method: The distance between a spatial unit and a cell tower is computed using their centroids. It is important to clarify that this method has only been used by researchers in combination with grids, not census tracts (Peredo et al. 2017; Ahas et al. 2015). Given $\boldsymbol{v}$ as the set of neighbor Voronoi polygons of spatial unit u, n_c as the footfall for cell tower c, and $d(u, v_c)$ as the distance between the centroids of u and v_c, the interpolated population for u is computed as follows:

$$n^*_{u,i=\text{Idw}} = \frac{\sum_{v_c \in \boldsymbol{v}} \frac{1}{d(u,v_c)} n_c}{\sum_{v_c \in \boldsymbol{v}} \frac{1}{d(u,v_c)}} \tag{4}$$

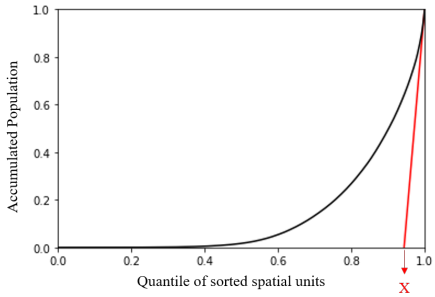
Let $\boldsymbol{s}^{(a,b)}$ be the set of spatial units and $\boldsymbol{v}^{(a,b)}$ be the set of Voronoi polygons intersecting with all spatial units in city a under boundary setting b. Interpolating using inverse distance weighting does not guarantee the sum of all $n^*_{u,a=\text{Idw}}$ for $u \in \boldsymbol{s}^{(a,b)}$ to be the same as the sum of all n_c for $v_c \in \boldsymbol{v}^{(a,b)}$. Therefore, we re-scale $n^*_{u,a=\text{Idw}}$ as follows:

$$n_{u,i=\text{Idw}} = n^*_{u,i=\text{Idw}} \frac{\sum_{v_c \in \boldsymbol{v}^{(a,b)}} n_c}{\sum_{u \in \boldsymbol{s}^{(a,b)}} n^*_{u,i=\text{Idw}}} \tag{5}$$

Hotspot detection

Given a CDR dataset, the *Loubar* method will be applied as follows. We will first aggregate the number of unique users for each cell tower c to obtain the average hourly number of unique users: $\overline{\boldsymbol{n}}_c = \{\overline{n}_{c,h}\}_{h=0}^{23}$, where $n_{c,h}$ represents the average of unique users between h:00:00 and h:59:59. Using equations (1) - (5), we can calculate the interpolated population with method i for a spatial unit u: $\overline{\boldsymbol{n}}_{u,i} = \{\overline{n}_{u,i,h}\}_{h=0}^{23}$.

Let $\boldsymbol{s}^{(a,b)}$ be all the spatial units in city a, $\boldsymbol{s}_m^{(a,b)}$ be the spatial units in a municipality m in city a in boundary setting b, and $\overline{\boldsymbol{N}}_{\boldsymbol{s},i,h} = \{\overline{n}_{u,i,h}\}_{u \in \boldsymbol{s}}$ be the interpolated population using method i for a set



Supp. Fig. 2. *Loubar* hotspot detection for a set of U spatial units with interpolated population. 1) the units are sorted in ascending order by the population; 2) draw the Lorenz curve of the accumulated population with x-axis being the ranking of units normalized by U; 3) compute the intersection of the tangent line at x=1.0 (the red line) and the x-axis. Let the intersection point be (X, 0); 4) The threshold δ is the population of the $U * (1 - X)$th spatial unit; 5) all spatial units with population $\geq \delta$ are the hotspots detected. The detailed explanation of *Loubar* method can be found in Louail et al. (2014).

of spatial units $\boldsymbol{s}$ at hour h. We apply the *Loubar* hotspot detection method to $\overline{\boldsymbol{N}}_{\boldsymbol{s}^{(a,b)},i,h}$ if b is Metro-UR or Metro-U, or apply *Loubar* to $\overline{\boldsymbol{N}}_{\boldsymbol{s}_m^{(a,b)},i,h}$ for each municipality m in city a if b is PerMuni-UR or PerMuni-U to compute the threshold value δ and detect whether a spatial unit u is a hotspot at hour h. Finally, a spatial unit will be identified as a hotspot if it is permanent *i.e.,* it is considered a hotspot throughout the 24 hours of the day (*all-day*) $\mathbf{1}_{u,i,h} = 1$ for $0 \leq h \leq 23$ (Louail et al. 2014). This binary decision can be denoted as:

$$\mathbf{1}_{u,i,h} = \begin{cases} 1, \text{if } \overline{n}_{u,i,h} \geq \delta \\ 0, \text{if } \overline{n}_{u,i,h} < \delta \end{cases} \tag{6}$$

However, since population density (Le Néchet 2012) and employment density (Giuliano and Small 1991) are often used in quantitative geography, which roughly correspond to the presence of people during nighttime and daytime, in this paper we will also explore *home-hour* and *work-hour* hotspots. These are formally defined as permanent hotspots during working hours (9am-5pm) or home hours (10pm-5am) *i.e.,* $\mathbf{1}_{u,i,h} = 1$ for $h_s \leq h \leq h_e$ with $h_s = 9$ and $h_e = 17$ and with $h_s = 22$ and $h_e = 5$, respectively.

Hotspot measurement variables

(1) Hotspot Scale quantified in terms of number and the total geographical area covered by the hotspots detected:

- **Number of hotspots (NHS)** is the number of spatial units that are detected as hotspots. Let hs be the set of hotspots and $|hs|$ be the number of hotspots:

$$\text{NHS} = |hs| \tag{7}$$

- **Area of hotspots (AHS)** is the total geographical area of the spatial units that are detected as hotspots. Let hs be the set of hotspots:

$$\text{AHS} = \sum_{u \in hs} \text{Area}(u) \tag{8}$$

(2) Urban sprawl

- **Compacity coefficient (COMP)** (Louail et al. 2014) measures the sprawl of the detected hotspots over a city, with smaller COMP values associated to less dispersed hotspots with respect to the size of the city. Let hs be the set of hotspots, $|hs|$ be the number of hotspots and $d_{j,k}$ the distance between the centroids of spatial units u_j and u_k, COMP is computed as:

$$\text{COMP} = \frac{\text{D}_{hs}}{\sqrt{\text{Area}(r_{a,b})}}, \text{D}_{hs} = \frac{\sum_{j=1}^{|hs|} \sum_{k=j+1}^{|hs|} d_{j,k}}{|hs|\,(|hs|-1)/2} \tag{9}$$

- **Mass Compacity coefficient (MCOMP)** (Le Néchet 2012) is a modified compacity coefficient that weights the distance between hotspots by the population of each spatial unit, and measures the average distance between individuals located within the detected hotspots. The smaller MCOMP is, the less dispersed the hotspots are with respect to the size and population of the city. Let p_j be the population in spatial unit u_j, MCOMP is computed as follows:

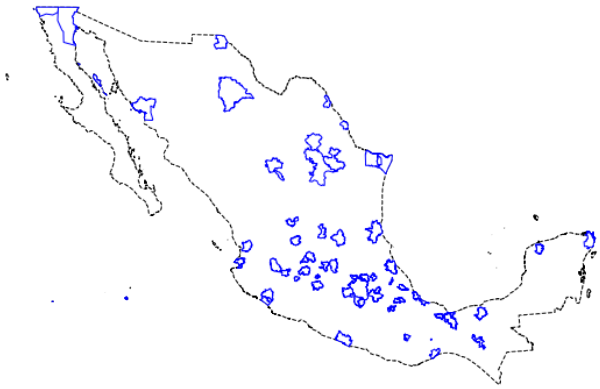
$$\text{MCOMP} = \frac{\text{MD}_{hs}}{\sqrt{\text{Area}(r_{a,b})}}, \text{MD}_{hs} = \frac{\sum_{j=1}^{|hs|} \sum_{k=j+1}^{|hs|} d_{j,k}\, p_j\, p_k}{\sum_{j=1}^{|hs|} \sum_{k=j+1}^{|hs|} p_j\, p_k} \tag{10}$$

(3) Urban compactness

- **Cohesion (COHE)** (Angel et al. 2010) is the ratio of the average distance-squared among all points in the reference circle and the average distance-squared among all points in the hotspot areas. Large cohesion means people in hotspot areas are very close to each other. Let hs be the set of hotspots, AHS be the total geographic area of hs, p be the points of hs in the rasterized format, $|p|$ be the number of points and $d_{i,j}$ be the distance between the i- and j-th point. The average distance-squared among all points in the reference circle $\approx \text{AHS}/\pi$. Finally, COHE is calculated as:

$$\text{COHE} = \frac{\text{AHS}/\pi}{\frac{2}{|p|(|p|-1)} \sum_{j=1}^{|p|} \sum_{k=j+1}^{|p|} d_{i,j}^2} \tag{11}$$

- **Proximity (PROX)** (Angel et al. 2010) is the ratio of the average distance from all points in the reference circle to its centre and the average distance to the geometry center of the hotspot areas. The proximity index focuses on the distance between points from the geometry center instead of the point-wise distance in the cohesion index. Let hs be the set of hotspots, g be the center of gravity of hs, AHS be the total geographic area of hs, p be the points of hs in the rasterized format, $|p|$ be the number of points and $d_{i,j}$ be the distance between the i-th point and the center g. The average distance of all points to the center in the reference circle $= \frac{2}{3}\sqrt{\text{AHS}/\pi}$. Thus, PROX



Supp. Fig. 3. 59 metropolitan areas in Mexico

is computed as:

$$\text{PROX} = \frac{\frac{2}{3}\sqrt{\text{AHS}/\pi}}{\frac{1}{|p|}\sum_{i=1}^{|p|} d_i} \tag{12}$$

- **Normalized moment of inertia (NMI)** (Li et al. 2013) is based on the dispersion of points from the center of its shape. It involves the calculation of the second moment of an area about a point, also known as the moment of inertia (MI). The MI is then normalized by the MI of the reference circle, hence normalized moment of inertia.
- **Normalized mass moment of inertia (NMMI)** (Li et al. 2014) takes into account the mass distribution of a shape. The previous three compactness indices consider only the geometric shape *i.e.,* each point in the shape is equally important in the compactness. Nevertheless, in our case, each hotspot might have a different estimated population or mass, and they can still be compact - even though their geometry shape is not - by having the majority of the population concentrate around the mass center. The reference circle in NMMI is no longer an equal-area circle, but a circle with equal-effective-area. The calculation of NMI and NMMI are rather long. The mathematical derivation can be found in (Li et al. 2013) and Li et al. (2014), respectively.

Results

Intra-city level analysis

The stability scores at the intra-city level for each hotspot index and across boundary settings are shown in Supp. Fig. 5(a). Standard deviation values for each stability score - representing the spread of the correlations among different combinations of spatial units and interpolation methods - are also shown. From the Figure, we can observe that:

1) Hotspot scale and urban sprawl indices are more stable in Metro-based settings, while urban compactness indices highest stability is achieved when using municipalities to define city boundaries.

	CT Uni	CT Pop	G Idw	G Uni	G Pop	Vor Uni	Vor Pop	Mean
CT Uni	1.00	0.55	0.47	0.50	0.65	0.15	0.23	0.43
CT Pop	0.55	1.00	0.40	0.49	0.68	0.30	0.31	0.45
G Idw	0.47	0.40	1.00	0.31	0.53	0.32	0.40	0.41
G Uni	0.50	0.49	0.31	1.00	0.57	0.18	0.06	0.35
G Pop	0.65	0.68	0.53	0.57	1.00	0.20	0.31	0.49
Vor Uni	0.15	0.30	0.32	0.18	0.20	1.00	0.80	0.32
Vor Pop	0.23	0.31	0.40	0.06	0.31	0.80	1.00	0.35

Supp. Fig. 4. Spearman correlation coefficient $Coef_{ind=PROX,b=Metro\text{-}UR,j,k}$ between each pair of combinations (C_j, C_k) for all-day permanent hotspots. The coefficient matrix is symmetric. The row mean is the average of coefficients in each row, excluding the values on the diagonal. The row mean of j-th row shows the average coefficients of combination C_j correlated with other combinations.

	Metro UR	Metro U	PerMuni UR	PerMuni U
NHS	0.38 (0.17)	0.38 (0.17)	0.31 (0.12)	0.35 (0.17)
AHS	0.43 (0.18)	0.41 (0.19)	0.36 (0.16)	0.39 (0.18)
COMP	0.52 (0.23)	0.52 (0.24)	0.30 (0.13)	0.33 (0.14)
MCOMP	0.49 (0.22)	0.49 (0.24)	0.36 (0.19)	0.40 (0.21)
COHE	0.25 (0.10)	0.29 (0.11)	0.29 (0.15)	0.35 (0.16)
PROX	0.26 (0.11)	0.30 (0.12)	0.29 (0.15)	0.35 (0.17)
NMI	0.25 (0.10)	0.29 (0.11)	0.29 (0.15)	0.35 (0.16)
NMMI	0.30 (0.11)	0.33 (0.14)	0.35 (0.15)	0.39 (0.15)

(a) 24-hour vector

	Metro UR	Metro U	PerMuni UR	PerMuni U
NHS	0.40 (0.14)	0.42 (0.15)	0.40 (0.16)	0.43 (0.18)
AHS	0.39 (0.15)	0.42 (0.16)	0.34 (0.14)	0.41 (0.17)
COMP	0.48 (0.21)	0.49 (0.23)	0.30 (0.12)	0.36 (0.15)
MCOMP	0.46 (0.22)	0.47 (0.23)	0.39 (0.18)	0.41 (0.22)
COHE	0.32 (0.14)	0.35 (0.14)	0.32 (0.16)	0.37 (0.16)
PROX	0.33 (0.15)	0.37 (0.15)	0.32 (0.17)	0.38 (0.16)
NMI	0.32 (0.14)	0.35 (0.14)	0.32 (0.16)	0.37 (0.16)
NMMI	0.37 (0.15)	0.39 (0.17)	0.39 (0.17)	0.44 (0.16)

(b) 4-hour-bin vector

Supp. Fig. 5. Stability (Standard deviation) of all indices in different boundary settings. The coefficients in (a) are computed using 24-hour vector of indices and in (b) are computed using 4-hour-bin vector. The gradient background color is based on the stability score ranging from 0 to 1, the darker the orange color is, the closer it is to 1.

Nevertheless, the impact of boundary settings is small for urban compactness indices. Based on these findings, researchers could favour one type of indices versus others depending on the boundary settings of interest, which would in turn produce more robust indices to measure hotspot rankings over time for a given city.

2) The highest stability score is achieved by urban sprawl indices (COMP and MCOMP) under Metro-UR and Metro-U boundary settings. This is different from the inter-city level stability analyses where the best score was achieved with PerMuni-based settings. However, even the highest stability is only moderately stable. AHS also has moderate stability under Metro-UR and Metro-U boundary settings. The rest indices in all settings are weakly stable, much lower than the stability scores at the inter-city level, indicating that a change in the boundary or spatial unit choice could produce widely different results.

3) The intra-city stability scores per index are computed extracting permanent hourly hotspots *i.e.,* hotspots that are considered as such throughout the 24 hours. As a result, the low stability scores could be potentially due to the stringent definition of hotspot. To assess that, we grouped the 24 hours into

6 bins - each bin is a 4-hour-bin - and computed the hotspot indices again. In this case, the Spearman correlation was computed between two 6-bin vectors of coefficients - instead of the previous 24-hour vectors. Supp. Fig. 5(b) shows that although the stability scores increase, except for the COMP index in the Metro-based setting, they are still mostly weakly stable.

Finally, it is important to mention that although the index stability scores at the intra-city scale are on average low, some cities do have much higher stability scores than others. This finding might be indicating that, unlike inter-city scales, intra-city hotspots' stability might depend on other types of cultural or social trends not studied in this paper.

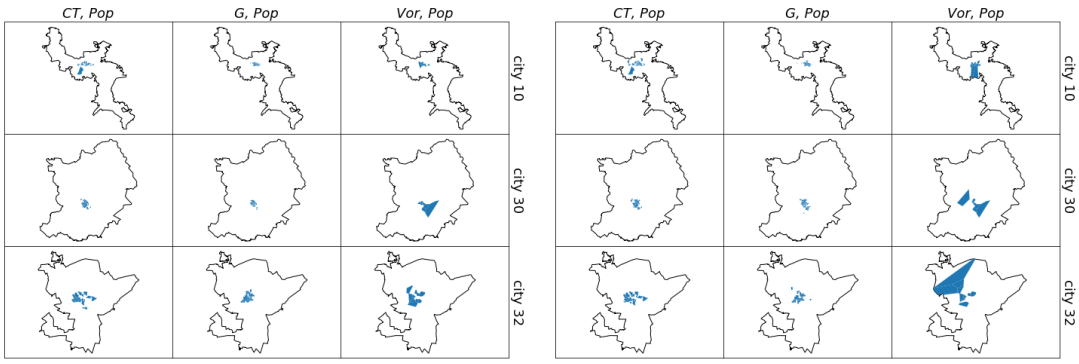
Discussion

Stability of urban compactness indices (COHE, PROX, NMI and NMMI) at the inter-city level

Urban compactness indices, similarly to urban sprawl, are computed using the distance between the detected hotspots. Therefore, these indices are subject to the same instability issues due to the various ways in which population and footfall can be distributed across municipalities, and to the varying shapes that Voronoi polygons might have in rural areas. Comparing Figures 3(e-h) and 3(i-l) (in the paper) we can also observe that the correlation between *Vor*-based and other spatial combinations is much worse for urban compactness indices than for urban sprawl indices. This is because compactness indices measure the compactness of the shape of hotspots, and the varying nature of the Voronoi polygons causes the *Vor*-based combinations to be weakly to no correlated with other combinations when the rural areas are considered.

Difference of stability between home-hour and work-hour permanent hotspots at inter-city level

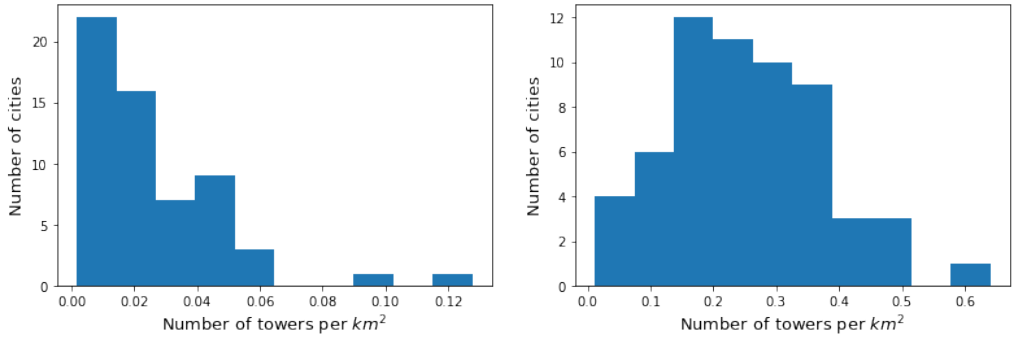
There exists little difference in the stability of hotspot indices computed for home- and work-hour periods across different city boundary settings. Nevertheless, it is worth noting that the stability scores for COHE, PROX and NMI computed for home-hour periods are slightly more unstable than their work-hour counterpart, especially in the Metro-UR setting. We argue that this might be due to the fact that for some cities home locations are more spatially scattered, possibly including the outskirts where CT and Vor polygons tend to be larger when compared to work locations in downtown areas that tend to have smaller CT and Voronoi polygons. Thus, these varying area sizes in home location polygons might be increasing the instability of the indices. See Supp. Fig. 6 for an example of this setting with city 10 (Tuxtla Gutiérrez), city 30 (Tepic) and city 32 (Oaxaca). The home-hour permanent hotspots are more scattered than the work-hour permanent hotspots. Using *(G, Pop)*, a few small grids away from the core area in city 30 and 32 are considered as permanent hotspots. But using *(Vor, Pop)*, the Voronoi polygons away from the core area have large variation in size. The variation in size would have a huge impact on the covered geographic area and subsequently on the equal-area circle considered in the compactness indices. Therefore, COHE, PROX and NMI have slightly lower stability in home-hour period.



(a) Work-hour permanent hotspots

(b) Home-hour permanent hotspots

Supp. Fig. 6. Permanent hotspots detected by methods *(CT, Pop)*, *(G, Pop)*, *(Vor, Pop)* for city 10, city 30 and city 32 in work and home hours under boundary setting b=Metro-UR



(a) Considering both urban and rural areas.

(b) Considering urban areas only.

Supp. Fig. 7. Distribution of cell tower density over all metropolitan areas under different boundary settings.

Relationship between stability and cellular tower density at the inter-city level

The cellular towers that collect CDR data are not homogeneously distributed across space. Instead, urban areas tend to have higher cellular density - more cellular towers per km^2 - than rural areas due to the size of the population they serve. Supp. Fig. 7 shows the cellular tower density, computed as the number of cell towers per km^2, across the urban and rural areas analyzed in this paper. We can in fact observe that the cell tower density is much larger in urban areas. In this section, we explore whether the density of cellular towers in spatial units is correlated to the variation of hotspot index measures computed under different methodological choices, which would in turn affect the hotspot index stability. If correlated, we could hypothesize that the hotspot index instability revealed for certain methodological combinations

	Metro-UR	Metro-U	PerMuni-UR	PerMuni-U
NHS	-0.50 ***	-0.07	-0.00	-0.02
AHS	-0.28 *	0.21	-0.12	0.25
COMP	-0.18	-0.31 *	-0.18	-0.10
MCOMP	-0.15	-0.28 *	-0.27 *	-0.30 *
COHE	-0.37 **	-0.38 **	-0.71 ***	-0.17
PROX	-0.29 *	-0.22	-0.65 ***	0.06
NMI	-0.37 **	-0.38 **	-0.71 ***	-0.17
NMMI	-0.42 ***	-0.39 **	-0.56 ***	-0.14

Supp. Table 1. Spearman correlation between variance of hotspot indices computed for different combinations (of spatial units and interpolation methods) and tower density of metropolitan areas. Significance levels: 0 '***'; 0.001 '**'; 0.01 '*'; 0.05 ' '

might be partially due to the cellular tower density, since high variance in index measures leads to high instability (diverse city rankings across settings).

To carry out this analysis, we propose the following approach. Given a boundary setting (Metro-(U)R, PerMuni-(U)R) and a hotspot index (NHS, AHS, COMP, MCOMP, COHE, PROX, NMI, NMMI), we compute the index value for each of the 7 combinations of spatial unit and interpolation method ((CT,Uni), (CT, Pop), (G, Uni), (G, Pop), (G, Idw), (Vor, Uni), (Vor, Pop)) and for each of the 59 metropolitan areas in our dataset. In other words, given a pair of boundary setting and hotspot index, we will have 59 vectors - one per metropolitan area - each containing 7 values, one per combination of spatial unit and interpolation method. Next, we compute the standard deviation across the 7 combinations for each of the metropolitan areas to measure the variation of a hotspot index; and we perform the Spearman correlation between cell tower density and hotspot index variation across all 59 metropolitan areas. For hotspot indices computed under boundary settings Metro-UR and PerMuni-UR, we use cell tower density considering both urban and rural areas; for hotspot indices computed under boundary setting Metro-U and PerMuni-U, we use density considering only urban areas. The correlation results are shown in Supp. Table 1.

The Table shows that most of the index variance is negatively correlated with cell tower density (except for three cases); and that statistical significance is high in metro settings (Metro-U and Metro-UR), settings that combine both urban and rural spatial units (Metro-UR and PerMuni-UR) and for all compactness indices. Interestingly, the correlations were not significant in the PerMuni-U setting (except for one value), which is the setting we have identified as providing the most stable hotspot indices. Overall, these results reveal that the higher the cell tower density, the lower the index variance, thus increasing index stability. Given that the correlation coefficients reflect a weak to moderate correlation, it is fair to say that part of the instability of the hotpot indices reported in the paper, might be due to the coarse-grained spatial information computed in spatial units with low cellular tower densities.

References

Ahas R, Aasa A, Yuan Y, Raubal M, Smoreda Z, Liu Y, Ziemlicki C, Tiru M and Zook M (2015) Everyday space–time geographies: using mobile phone-based sensor data to monitor urban activity in harbin, paris, and tallinn.

Int. J. Geogr. Inf. Sci. 29(11): 2017–2039.

Angel S, Parent J and Civco DL (2010) Ten compactness properties of circles: measuring shape in geography. *The Canadian Geographer / Le Géographe canadien* 54(4): 441–461.

Giuliano G and Small KA (1991) Subcenters in the los angeles region. *Reg. Sci. Urban Econ.* 21(2): 163–182.

Le Néchet F (2012) Urban spatial structure, daily mobility and energy consumption: a study of 34 european cities. *Cybergeo* .

Li W, Chen T, Wentz EA and Fan C (2014) NMMI: A mass compactness measure for spatial pattern analysis of areal features. *Ann. Assoc. Am. Geogr.* 104(6): 1116–1133.

Li W, Goodchild MF and Church R (2013) An efficient measure of compactness for two-dimensional shapes and its application in regionalization problems. *Int. J. Geogr. Inf. Sci.* 27(6): 1227–1250.

Louail T, Lenormand M, Cantu Ros OG, Picornell M, Herranz R, Frias-Martinez E, Ramasco JJ and Barthelemy M (2014) From mobile phone data to the spatial structure of cities. *Sci. Rep.* 4: 5276.

Peredo OF, García JA, Stuven R and Ortiz JM (2017) Urban dynamic estimation using mobile phone logs and locally varying anisotropy. In: Gómez-Hernández JJ, Rodrigo-Ilarri J, Rodrigo-Clavero ME, Cassiraga E and Vargas-Guzmán JA (eds.) *Geostatistics Valencia 2016*. Cham: Springer International Publishing. ISBN 9783319468198, pp. 949–964.

Statstutor (2020) Spearman's correlation. `http://www.statstutor.ac.uk/resources/uploaded/spearmans.pdf`. Accessed: 2020-1-17.

Xu W, Chen H, Frias-Martinez E, Cebrian M and Li X (2019) The inverted u-shaped effect of urban hotspots spatial compactness on urban economic growth. *R Soc Open Sci* 6(11): 181640.