SPATIAL COGNITION AND COMPUTATION, vol(issue), start–end

Investigating Behavioural and Computational Approaches for Defining Imprecise Regions

Florian A. Twaroch[1], Paul Brindley[2], Paul D. Clough[2], Christopher B. Jones[1], Robert C. Pasley[3] and Sue Mansbridge[2]
Cardiff University, UK[1]
University of Sheffield, UK[2]
University of Nottingham, UK[3]

People often communicate with reference to informally agreed places, such as 'the city centre'. However, views of the spatial extent of such areas may vary and result in imprecise regions. We compare perceptions of Sheffield's City Centre from a street survey (with 61 participants) to spatial extents derived from various web-based sources. Such automated approaches have advantages of speed, cost and repeatability. Our results show that footprints derived from web sources are often in concordance with models derived from more labour-intensive methods. There were, however, differences between some of the data sources, with those advertising/selling residential property diverging the most from the street survey data. Agreement between sources was measured by aggregating the web sources to identify locations of consensus.

Keywords: geographic information retrieval, place, spatial cognition, social media, vague geography.

Correspondence concerning this article should be addressed to Dr Paul Brindley, The University of Sheffield, Arts Tower, Western Bank, Sheffield, S10 2TN, UK; email: p.brindley@sheffield.ac.uk.

Introduction

People often refer to place in daily communication; however they can have different views of where a place actually is. Where this vernacular language is used to refer to places that have the same names as administrative ones, the spatial interpretation often does not correspond to the formal spatial definition. It has been observed that most geographic objects seem in some way to be an abstraction of things and have unclear and fuzzy boundaries; whilst relatively few objects in geographic space have precise boundaries (Burrough, 1996; Frank, 1996; Fisher, 2000; Couclelis, 2003). Thurstain-Godwin & Unwin (2000) describe city centres as "... almost archetypal examples of geographic objects with indeterminate boundaries ...". Other authors, such as Bettencourt (2013) and Masucci et al. (2015), view cities as a special form of place, poorly defined and dependent on various attributes, both physical and notional.

The source of vagueness can be due to imperfections in the observation process, but is inherent to many geographic objects. People perceive and represent spatial reality in an internal cognitive model, but when spatial experience is communicated to others a description language is used to externalize and transfer a representation of this internal cognitive model (Glasersfeld, 1996). These individual cognitive internal models might be expected to coincide with other people's models.

Research into the cognition of vernacular regions has received attention from the developers of tools and services that provide access to geographical information (Schockaert et al., 2005; Purves et al., 2005; Arampatzis et al., 2006; Jones et al., 2008). For example, the usability of Geographic Information Retrieval (GIR) systems can be improved by dealing effectively with vague and vernacular information (Goodchild, 2000; Vögele et al., 2003; Purves et al., 2007; Schockaert, 2011). References to places, such as the 'City Centre', are also frequently used in web searches or calls to emergency services and could significantly improve the quality of such information services (Davies et al., 2009). Therefore methods to collect and represent informal place names and build up spatial representations are necessary.

In previous work on the definition of the vague region 'downtown' Santa Barbara, Montello et al. (2003) came to the conclusion that people readily provided information about the spatial extent of a vague region if given the appropriate map media. They observed a high degree of agreement between respondents' definitions. In discussing the results they posed the question of whether an effect would have been observed if they interviewed subjects by different means. In this paper, the approach proposed by Montello et al. (2003) has been modified to use perceptions of landmark membership rather than maps in the questionnaire. We investigate people's perception of Sheffield City Centre in the UK and attempt to establish whether computational techniques utilising web-based sources of data are able to produce comparable results. The advantages of such automated methods include reducing the time required to acquire region footprints and improved repeatability. It would thus enable researchers and professionals to obtain knowledge of larger numbers of places, possibly with finer detail. Specifically the research questions addressed in this study are the following:

- [RQ1]: To what extent do people agree on the 'membership' of landmarks within Sheffield City Centre and does the location at which a person is interviewed and their familiarity with an area affect their perception of the 'City Centre'?

- [RQ2]: To what extent do the representations of 'City Centre' derived using data collected manually differ from or agree with those automatically gathered from web-based sources?

Our work contributes to the growing body of literature utilising georeferenced data extracted from online sources to characterise geographic regions. The novelty of our work includes a comparison between the geometric footprints for city centres derived from six web-based sources with boundaries produced using more labour-intensive manual data collection methods. To date, no previous studies have compared vague representations <u>of a city centre</u> across such a range of different data sources (though Gao et al. (2017) used five web sources in a study of the regions of Northern and Southern California).

The remainder of the paper is structured as follows: After a short definition of the terminology used in our study we review related work on the representation of vaguely cognised regions. Following this, we explain our experimental setup and analyse the collected data. To investigate the research questions, we conducted a street survey with pedestrians in Sheffield City Centre (hereafter referred to as the s*treet survey*) to obtain people's feedback on the membership of landmarks to the City Centre. We use Fleiss' Kappa statistic to compare the agreement of the answers given by subjects at three different locations. We create geometric extents from subjects' responses using a Kernel Density Estimation (KDE) technique (Silverman, 1986). Multiple thresholds of the KDE surfaces are used to test similarity between interview locations and to evaluate the automated methods. The continuous KDE surfaces derived from different data sources are compared using linear regression to draw conclusions about the different representations for Sheffield City Centre. Precision and Recall measures are applied to multiple thresholds of the KDE surfaces to compare the output with existing crisp definitions of Sheffield's City Centre, based on previous academic research (Lüscher & Weibel, 2013) and Sheffield City Council definitions (City Alert Scheme). Results show agreement between both the manual and automated methods used in this work to represent Sheffield City Centre and existing definitions from previous work. We also investigate whether the location of interview plays a role in people's perception of Sheffield City

Centre and whether answers are influenced by the subject's familiarity with the environment.

Perception, Cognition and Representation of Spatial Regions

Perception and cognition of spatial reality underlie complex processes that are shaped from early childhood (Piaget & Inhelder, 1948; Spelke, 1990). People's familiarity and experience with a spatial environment play a major role in defining representations of spatial reality. The study by Montello et al. (2003) investigated individual perception of the location and extent of downtown Santa Barbara. The study used a questionnaire-type method with three tasks: participants were asked to draw an outline of the downtown area on a base map of the wider region; they were then asked to repeat the task but with 50% and 100% confidence that the outline enclosed the area; finally participants were asked to mark the 'core' of the downtown area. A total of 36 pedestrians were involved in the final results based on interview at eleven different locations. In a discussion of the results by Montello et al. (2003), they questioned if their methodology was biased by the base maps that had been used. In a later study involving larger regions, Montello et al. (2014) overlaid a grid of hexagonal cells over computerised maps. This allowed them to gather vague perceptions of region membership from each respondent rather than averaging over several respondents to produce a vague repesentation of a region.

In contrast to these labour-intensive manual studies, efforts have been undertaken to automate, to some degree, the process of representing place. Automated definition of vaguely cognised places in the UK has been based on census and socioeconomic data (Thurstain-Goodwin & Unwin, 2000), web map tools (Evans & Waters, 2007) and analysis of landscape features (Fisher, 2004). The widespread availability of *volunteered geographic information* (Goodchild, 2007) has also led to approaches to mine such web-based sources in order to build representations of vaguely cognised regions. For example, a number of researchers, (notably Purves et al., 2005; Arampatzis et al., 2006; Schockaert et al., 2005; Jones et al., 2008) have used data from search engines to automatically extract information to model the spatial extent of place names based on the locations of associated place names, found with named entity recognition methods. The associated places were georeferenced with gazetteers such as the Alexandria Digital Library (Hill et al., 1999) and the Getty Thesaurus of Geographic Names (Harpring, 1997). The relevant web pages were found using a variety of types of queries that named the target place to be modelled. Some of these queries included phrases that expressed a spatial relationship between the target place and found georeferenced places. The spatial extents were modelled using kernel density estimation (e.g. Purves et al., 2005), Delaunay triangulation related methods (Arampatzis et al., 2006) and fuzzy modelling (Schockaert et al., 2005). Fuzzy modelling was also used in Schockaert & De Cock (2007) to represent neighbourhoods based on data points for businesses retrieved for each named place with a local search API. Gao et al. (2017) used a cell-based data-synthesis-driven method for detecting and extracting vague vernacular regions. They automated an earlier manual method described in Montello et al. (2014) using data from social media postings. They found that the automatic methods produced results that correlated significantly with the manual methods. At the same time these methods have the advantages that they can be repeated and different scales can be used without the limitations of using human participants. The regions used in that study – Southern California and Northern California – are large in comparison with city centres. It was found that people's perceptions of these regions did not cover the south and north respectively of the state of California by any means or even reach to the most northerly and most southerly points of the state.

Other researchers used georeferenced content, such as images from the online photo sharing website Flickr.com, to create representations of boundaries (Hollenstein & Purves, 2010; Mackaness & Chaudhry, 2013; Chen & Shaw, 2016). Lüscher & Weibel (2013) used a survey to gather typicality measures for various urban features. They used rich data sets containing such features to define city centre boundaries for a set of UK cities. Their results yield plausible boundaries which compare well to other definitions, as well as those of Hollenstein & Purves (2010).

Investigations on how well these automated methods reflect peoples' cognitive models of vaguely defined places, however, are still rare. Tezuka & Tanaka (2005) measured the cognitive significance of landmarks based on web counts, linguistic analyses and proximity measures. The authors point out that landmarks that are visually significant objects are not always the objects frequently referred to by people. This is reflected in the frequency of used landmark terms in web documents. In the study, Tezuka & Tanaka (2005) asked 50 subjects to name the 20 most notable landmarks in Kyoto. They then compared these human judged landmarks to landmarks extracted from web documents and concluded that measures that considered spatial context performed better than classical document and term frequency measures in predicting the use by people of these landmarks as reference points. Our work compares behavioural and computational approaches to provide greater insights on how well automated methods utilising multiple web-based sources of data can be used to derive geometric footprints for imprecise regions.

Method

Initially a street survey was undertaken to ascertain differing perceptions of Sheffield's City Centre. Vague boundaries were then generated using automated methods from web-based data from six different data sources. This enabled a comparison to be made to assess how well these automated methods reflect people's cognitive models of Sheffield's City Centre.

Street Survey

The investigation of individuals' perceptions of the extent of Sheffield City Centre used a street survey that involved interviewing pedestrians as respondents. This differs from the experimental setup of Montello et al. (2003) in two ways. Firstly, no maps were used; subjects were asked to judge the 'membership' of 38 landmarks in Central Sheffield. They were given the possibility to express no knowledge of the location of a landmark. Respondents were not led by any visual aids, to avoid any bias that might be inherent in the use of maps. However, other biases may have existed, such as familiarity with the landmark or the city in general, the choice by the researchers of suitable landmarks, the sparsity of landmarks in some areas, the use of linear features such as streets, unnamed points, and lack of provenance for some web sources. Subjects were not asked to draw confidence boundaries; rather these were calculated in a subsequent step from the given point data based on membership of landmarks. A second difference was that the surveys were conducted at just three different locations in central Sheffield (Montello et al. (2003) used eleven) to enable investigation of whether peoples' perceptions of City Centre were affected by their current location.

Landmarks used in the study were chosen after an intensive study of six different schemes used to define the City Centre: Sheffield City Centre Alert Scheme, Open Forum for Economic Regeneration, City Centre Management, Council Planning Department, Council Tax Department, and local Emergency Services. Additionally, a pilot study was conducted to assess the suitability of selected landmarks for the questionnaire (see below). Landmarks that lie within and outside the previous six definitions of the City Centre were selected. Figure 1 shows the location of the chosen landmarks and the three locations used to interview people: The Moor, Peace Gardens and the Railway Station. In addition, we performed an automated comparison of the outcomes of the street survey to representations generated with automated techniques.

A pilot study was carried out to validate the street survey. The time taken to complete the survey ranged between five and ten minutes. The instructions and explanations were changed slightly in response to comments from the pilot. For example, respondents' embarrassment at lack of knowledge led to a greater emphasis on explaining that the City Centre does not have an unequivocal definition and that all answers were acceptable. Also, places that are clearly well-known to a participant were not necessarily known by name. For example, many respondents gave shopping as one of their reasons for visiting Sheffield City Centre but some said they did not know the location of Fargate, despite this being one of the major shopping areas in the city.

People were approached at three different locations across Central Sheffield (see Figure 1). The locations were chosen to be representative of typical Central Sheffield locations and to allow us to safely stop passers-by and interview them. They are busy areas in the city and thereby increased the number of potential participants. People had the option to participate or to decline to take part in the survey. A total of 65 people (37 females and 28 males) agreed to participate. The results of four subjects were subsequently excluded from the survey because they failed to complete the task in full, leaving 61 participants. 18 (30%) participants were aged less than 26; 9 (15%) aged 26-35; 8 (13%) aged 36-45; 7 (11%) aged 46-55; 8 (13%) aged 56-65 and 11 (18%) aged over 65.

Table 1 shows the locations and number of respondents interviewed. Prior to conducting the survey, a series of landmarks was compiled to include those within, outside and on the boundary of Sheffield's administratively-defined City Centre. They were based on the existing administrative definitions of Sheffield City Centre, as indicated above. However, there are parts of this area that present no obvious landmarks, for example the predominantly residential north west section of the central area. In the pilot test of the questionnaire this area had no landmarks, but subsequently one was added which it was judged might be known (Shalesmoor), as a stop for the city tram system bears the same name.

Several landmarks were included because they were within an administrative definition of the City Centre and would be well known and therefore likely easier for respondents to comment upon. It was felt that the list should include some very well-known landmarks to allow the participants some relatively easy decisions. These included the Town Hall, City Hall, the Crucible Theatre, the Lyceum Theatre and the Cathedral. Some landmarks were omitted because they might have introduced bias due to their names, for example the Central Library. Linear features, such as streets, were avoided in general since part, but not the whole length, might be perceived as being in the City Centre.

The labels of the landmarks were presented in alphabetical order on a paper questionnaire to the subjects. For each landmark four options were available as checkboxes: "inside", "outside", "on the boundary" and "don't know where it is". In the pilot study, a further option of "maybe" was included. This was because it seemed possible that people would make their answer dependent on the purpose of a visit of the City Centre. However, there was no evidence this option was used as expected, so it was removed for the main study. Subjects were also asked for their home postcodes and the duration they had lived and studied/worked in Sheffield. Subjects were asked for only the first part, the "postcode district"[1], of their postcode, which enabled an approximate geo-referencing of subjects' home locations.

Geographical representation of street survey output

Our method of modelling the extent of the city centre is based on the use of the Kernel Density Estimation method (KDE), which enables both visualisation of variation in confidence of the extent and the derivation of precise boundaries based on particular confidence levels. Each landmark was converted into point

[1] There are about 3,000 postcode districts in the UK, a full UK postcode represents on average 15 households

coordinates x_i and assigned a positive point weight w_i. The weight w_i has been calculated as the sum of inside points: "yes" (1) and "on the boundary" (0.5) answers r_{ik} from individual subjects k for the i^{th} landmark normalised by the total number of subjects n (eqn. 1). Outside points are not used in this weight.

$$w_i = \frac{\sum_{k=0}^{n} r_{ik}}{n} \qquad \text{(eqn.1)}$$

This weighted point set was then used to apply the KDE technique (Silverman, 1986) and to visualise the results on a map. A bandwidth (or search size) of 400 metres was chosen manually on the basis of adapting to the scale of the space being examined. This value has been used in other studies that create KDEs of urban data (e.g., Robinson et al., 2016) and it was used for all KDEs in this paper. Sensitivity analysis was undertaken using different bandwidths in order to ensure output and findings were robust. A grid resolution (cell size) of 50 metres was used throughout (as used in similar studies by Li & Goodchild, 2012; Hollenstein, 2008 and Brindley et al., 2017). The KDE method provides a tool to transform a set of points to a continuous surface representation that allows the density of the points to be estimated at any location. We do not normalise the KDE since we do not use it as a probability distribution. The principle of KDE is based on determining a weighted average of data points within a moving window centred on a grid of points p.

$$f(x) = \frac{1}{n \cdot h} \sum_{i=1}^{n} k\left(\frac{\| x - x_i \|}{h} \cdot w_i\right) \qquad \text{(eqn. 2)}$$

In the above equation (eqn. 2), n refers to the number of observed points x_i and k is the kernel function that is often unimodal and symmetrical (Brunsdon, 1995); x stands for any location in space and has to be interpreted as a vector, as well as x_i. The value w_i is the weight associated with each landmark as described above. The outcomes of the KDE are predominately influenced by the choice of the study region, the chosen bandwidth h parameter and the grid resolution, and less by the choice of the kernel function (Brunsdon, 1995; O'Sullivan & Unwin, 2002). If desired, thresholds can then be calculated at different levels for the resulting surface in order to gain crisp representations.

Computational Methods

In the light of developments in Geographic Information Retrieval for extending and enhancing geographic resources with data mined from the web, we compare the output from the manual street survey with models of Sheffield City Centre in the UK that are generated automatically using computational methods from various web-based sources - as described subsequently. The use of computational and web-based methods to model and analyse sociological and cultural phenomena is also becoming more commonplace, for example in the area of computational social science (Cioffi-Revilla, 2017). Some of the web data used here has been obtained from previous studies that included data for Sheffield, in particular Twaroch et al. (2008), who mined user contributed web data that provided the name ‘‘City Centre’’ for specific point locations, and Brindley et al. (2017) who mined web-based address data that contained references to “City Centre”. The web sources are:

- Georeferenced Flickr photographs that include the text “Sheffield City Centre” within the title, description or tags;
- Google business addresses that include “City Centre” (with location coordinates derived from the address postcodes and as used by Twaroch et al. (2008));
- Google community places (user contributed named places that include “City Centre’ with coordinates corresponding to the provided mapped locations and as used by Twaroch et al. (2008));
- Gumtree advertising web site (coordinates for the postcodes that contributors specify in association with a neighbourhood place name – i.e. “City Centre” and as used in the study by Twaroch et al. (2008)). Currently permission is required to mine Gumtree but the data used here were mined before such terms of use were in place;
- Rightmove estate agent (residential and non-residential) properties specified as being within the “City Centre.” Permission is required to automatically retrieve data; our dataset was gathered manually; and
- Web extraction – data containing references to the “City Centre” within Sheffield postal addresses found on the web (using the Bing web search API), where the geolocation was derived from the postcode (the method used is the same as described in Brindley et al. (2017)).

Geographical representation of results of computational methods

The results of mining data from these sources are point sets for which KDE can be applied as a means of generating a continuous surface. KDE was undertaken using the same approach as previously described with two exceptions. Firstly, in contrast to the point sets obtained in the street survey, the points mined from the web are not associated with weights. All mined points are interpreted as being “inside” the city centre region and unlike the survey data there are no points representing “on the boundary” or “outside.” The resulting KDE surface measures the spatial density of the distribution of these mined data points each of which is assigned a value of 1. Secondly, web data, such as described above, may contain elements of erroneous data. Most similar work uses the KDE surfaces to remove stray outlying points by discounting the lowest 10% (Hollenstein, 2008) or 20% of values (Twaroch et al. 2009).

There are, however, inherent difficulties in applying the same KDE method to remove outlying error within numerous datasets that themselves contain very different amounts of error. We found that in such circumstances, output across the data sources was most comparable when adopting an approach of requiring enough data points to give assurance in the output. After testing, we determined that at least five data points within the 400 metre bandwidth provided appropriate outputs (i.e., only including grid cells where the KDE surface was greater than 0.3 data points per hectare). Data with less than five data points within 400 metres were excluded from further analysis. Sensitivity analysis was also undertaken using different cut-off criteria (3 points within 400m: KDE400 > 0.18 data points per hectare, and 10 points within 400m: KDE400 > 0.6 data points per hectare) in order to ensure the output and findings were robust.

Comparison of geographic representations generated by the different approaches

Evaluation methods based on traditional Precision and Recall measures are not suitable for comparisons of *continuous* surfaces. Instead, linear regression (as used within Brindley, 2016) was undertaken in order to compare the KDE values for every grid cell between two given data sources. The process was repeated for each combination of the different data sources. Grid cells which were greater than zero in either of the comparator data sources were included within the regression output.

In contrast to comparisons of continuous surfaces, similarity of crisp boundaries can be assessed using traditional Precision and Recall approaches. Crisp definitions were generated from the KDE surfaces (as shown in Figure 2 within the results section) and compared with other crisp definitions of “City Centre”, such as the City Alert Scheme and as generated by Lüscher & Weibel (2013).

It is also possible to aggregate the KDE surfaces of the different computation data sources in order to establish the level of conformity. The number of different sources in agreement in each cell might reflect an overall level of consensus for the cell being called the “City Centre.” It may also be the case that the combined superset might more accurately reflect general opinion than any separate individual data source. This was achieved through converting each KDE surface into a binary version (with values greater than zero being assigned a value of one and all other cells being zero) and then totalling all the binary surfaces. The decile contours from this aggregation of the computational methods were also compared against the previously described existing crisp definitions and 50% contour from the street survey. The continuous grids were converted to a decile classification with an equal number of cells in each of the ten categorises. The cut-off thresholds between the decile groups were then used to form contours. For example, the following decile groups were generated from the KDE surface for the web extracted data (values represent the number of data points within 400m): 0-0.2; 0.2-0.5; 0.5-0.7; 0.7-0.9; 0.9-1.1; 1.1-1.4; 1.4-1.6;

1.6-1.8; 1.8-2.0. Contours were generated at the following KDE values: 0.2, 0.5, 0.7, 0.9, 1.1, 1.4, 1.6, 1.8.

A 50% contour from the street survey was used as it represented the decile contour with the closest similarity to the existing crisp definitions based on an average F_1-score measure. F_1-scores (along with precision and recall measures) were generated based on the number of cells within the boundary lines. Thus, it was possible to compare between decile contours and crisp boundary comparisons.

Results: Street Survey results

The role of location

Our questionnaire allowed several ways of recording the location of subjects and landmarks. Postcodes of participants' home addresses were associated with coordinates using the Ordnance Survey CodePoint (postcode locations) dataset. Landmarks used in the survey were geo-referenced by digitising them using an online mapping service (Google Maps). The distances of participants' home postcodes to the City Hall of Sheffield (a landmark all subjects knew and agreed to be part of the City Centre) were calculated, and the distributions of these distances for each of the three locations compared.

Results show that subjects interviewed at The Moor lived in a range of 2.5-18km from the City Centre (mean of 9km). Similarly, subjects interviewed at the Peace Gardens lived between 3-17km from the City Centre (mean of 8.7km). In contrast, however, participants who were interviewed at the Railway Station lived between 0.5-227km from the City Centre (mean of 21km). As one might expect subjects interviewed at the Railway Station tended to live further away. When considering all interview locations most people (42, 69%) lived within a range of 10km of the City Centre.

'Don't know where it is'

In total, 9 (15%) of the subjects (6 at The Moor, 2 at the Peace Gardens, 1 at the Railway Station) knew *all* of the landmarks in the questionnaire. All subjects knew the whereabouts of the City Hall and the Railway Station. Most of the subjects did not know at least one of the landmarks. Table 2 shows the distribution of "don't know" responses for all landmarks in the survey at the three locations. Notably, the total number of "don't know" responses at the Railway Station (22%) was higher than in the other two locations (13% at The Moor and 14% at the Peace Gardens). Questionnaires completed at the Railway Station were from people covering a range of familiarity with Sheffield; some knew it very well, having lived in the city for many years, while others knew it only slightly from a brief period of living in Sheffield or from visits from their homes in nearby places, such as Chesterfield and Derby.

Several of the landmarks seemed to be broadly unknown, at least by name, by participants. For some of these landmarks, for example Devonshire Green, the Fire/Police Museum and Mappin Street, the proportion of participants who answered “don’t know” was evenly spread across the three interview locations (see Table 2). However, for a number of landmarks with a high incidence of ‘don’t know’ answers, there were differences between the three locations (Table 2). There is little correlation (r=-0.15) between the distance of respondents’ home residence from the City Hall (a well-known central location) and the number of “don’t know” responses. This is a result of people working for a long period of time in Sheffield, but living outside of the city (e.g., one subject who has been working in Sheffield for 16 years reported that they lived in London, about 270 km away by road).

Membership of landmarks

We ranked the landmarks based on the number of inside and outside responses (see Table 3). This suggests that subjects do agree on certain landmarks being definitely inside and respectively outside the city centre. The top ranked “on the boundary” landmarks were Victoria Quays, Waitrose, Royal Victoria Hotel, the Wicker, Ponds Forge and the Railway Station. Overall, just over 11% of the respondents used the response “on the boundary” in describing the location of landmarks relative to Sheffield City Centre. We do not speculate whether this is due to few landmarks falling where people perceived the boundary to be or because the concept is not clear.

How much did participants agree?

We calculated the inter-rater agreement for each location using Fleiss’ Kappa, κ, Statistic (Fleiss, 1971). The measure allows the calculation of agreement between m raters on categorical data. It has been used for example in medicine and information retrieval (Fleiss, 1971; Carletta, 1996).

In our case, κ can be thought of as measuring how consistently subjects named a landmark as being “inside”, “outside” or “on the boundary” of the City Centre. Landis & Koch (1977) provide an interpretation of the κ scores (see Table 4). We used R (using the IRR package) to calculate Fleiss’s Kappa, marking ‘don’t know where it is’ responses as missing data (NA). According to the interpretation of the Fleiss κ statistic, the results show ‘moderate’ agreement among subjects at the Peace Gardens compared to ‘fair’ agreement for subjects interviewed at The Moor and the Railway Station. The agreement between respondents within each location is therefore fairly low. The overall agreement was ‘slight’.

We re-calculated the Fleiss κ statistic for subjects who had lived or worked 5 years or more in Sheffield. The results are shown in Table 5. We note there is less variation between the Kappa κ scores for the results of subjects who lived or worked ≥5 years in Sheffield. The interrater values are all fair to moderate, including the overall agreement.

Agreement between interview locations

We tested agreement between interview locations. We created a score for each location based on the responses. A response of “inside” was counted as 1, a response of “on the boundary” was counted as 0.5 while “outside” was counted as 0. We compared the locations using the Mann-Whitney U Test between pairs of the three locations. The results (Railway Station/Peace Gardens: U=702, p=0.83; Moor/Peace Gardens: U=571, p=0.12; Moor/Railway Station U=589, p=0.17) show that when the responses from the three separate locations are compared they show no significant difference between groups of respondents. There was however a low level of interrater agreement within the locations (see Table 5).

Results of the different geographic representations

Figure 2 shows the KDE surfaces as crisp definitions based on decile contours from the various data sources:

- Flickr image repository (Flickr) [number of data points (n) =522];
- Google business postcode addresses of businesses (Google BM) [n=405];
- Google community user-generated content (Google CM) [n=505];
- Gumtree website which consists of classified and real estate adverts (Gumtree) [n=114];
- Rightmove estate agent data (Rightmove) [n=250];
- Web extraction of address based information (web extraction) [n=288]; and
- Primary data collection from the manual street survey (street survey).

High threshold values in Figure 2, i.e. close to the maximum density value of the KDE, can be regarded as expressing a high assurance in the model. High values may result in several local maxima and therefore several contour polygons.

Comparisons of the different geographic representations

The R-squared values from the linear regression are shown in Table 6 and compare the KDE surfaces for each combination of data sources. These demonstrate the overall level of similarity between the continuous KDE surfaces for each data source. Diverse results were obtained when comparing the continuous KDE surfaces from the various computational methods with that generated from the street survey (Table 6). Whilst 76% and 73% of the variance in the street survey output could be explained by the web extraction and Google BM KDE surfaces respectively, only 48% and 1% of variance in the street survey KDE surface could be explained using the Rightmove and Gumtree representations.

Comparisons were also made against existing crisp definitions of "City Centre" (such as the City Alert Scheme and output from Lüscher & Weibel (2013), the boundaries of which are shown in Figure 3). The F_1-value from this traditional Precision and Recall evaluation can be found in Tables 7 and 8, whilst more comprehensive results can be found in Online Supplementary Tables 1 and 2.

Table 7 shows that there was a high degree of agreement between the City Alert Scheme definition and the web extraction and street survey representations using the 60% contour (F_1-scores of 86% and 85%). In contrast, the highest F_1-score for the City Alert Scheme boundary was 79% for the Google BM data; 73% for the Flickr representation; 71% for both Google CM and Rightmove data; and 37% for Gumtree information.

Broadly similar levels of agreement were found with the Lüscher & Weibel (2013) crisp definition (Table 8). The highest F_1-score was 89% for the web extraction approach; 87% for both Rightmove and street survey data; 79% for Google BM information; 73% for the Flickr representation; 66% using Google CM data; and 49% for Gumtree information. The main difference between the two different comparisons of crisp definition was that the Rightmove data was more closely aligned to the Lüscher & Weibel (2013) boundary.

The geographic extent of the combined/aggregated surface from the output of the computational methods is shown in Figure 3. The overall impression from Figure 3 is one of similarity. There appears, on face value, a reasonable comparison between the crisp definitions of Sheffield's City Centre and the geography from the aggregated computational methods (particularly when five or more of the six different sources were in agreement that the cell was within the City Centre).

The extent to which the different sources of information (Google, Gumtree, and so forth) contribute to the aggregated surface from the output of the computational methods (Figure 3) is shown in Online Supplementary Figure 1. This demonstrates that where not all six data sources were in agreement with the 50% contour from the street survey, it was most likely due to the absence of support from Gumtree.

Precision and Recall for the comparison of the aggregated output against existing crisp definitions can be found in Table 9. This shows that the highest F_1-score between the City Alert Scheme geography and that from the aggregated computational methods is 86% (when at least five of the different computational sources were in agreement). This is slightly higher than the highest F_1-score when comparing the City Alert Scheme against any single computational approach (F_1-score of 85% for the web extraction comparison, as shown in Table 7).

The geography for the aggregation of computational outputs (as shown within Table 9) however showed a lower level of agreement than did some single computational sources with the crisp definition from Lüscher & Weibel (2013) shown within Table 8, and the 50% contour from the street survey shown in Online Supplementary Table 3. Thus for both of these latter two measures, the output from the single computational sources from the web extraction, Rightmove and Google BM data each provided higher levels of agreement than the geography based on the aggregation of computational methods output.

This said, however, the differences between the existing sources of crisp definitions should be noted. Comparison between the City Alert Scheme and Lüscher & Weibel (2013) crisp definitions produced an F_1-Score of 74.5% (Precision: 59.4%; Recall: 100% - full result not shown). Because these crisp definitions differ in their areal extent then this has implications for the selection of contour values for the automated methods (i.e., contour values may end up fitting one of these definitions rather than both). However, when comparing the aggregated computational output against each of the three crisp definitions, the best fit with discrete geographies was obtained when at least five of the six sources were in agreement.

Sensitivity analysis

Sensitivity analysis was also undertaken to ensure robustness of findings. This consisted of (i) using different bandwidths within the KDE process (300m and 500m instead of 400m) and (ii) using different cut-off criteria for removing stray, erroneous data points (3 points within 400m and 10 points within 400m instead of 5 points within 400m as used within the main analysis). Using 3 data points within 400m would be the same as only including those cells in analysis where the KDE values were greater than 0.18 data points per hectare; whilst the use of 10 points within 400m is analogous to selecting only cells with KDE values greater than 0.6 data points per hectare.

Online Supplementary Tables 4 and 5 demonstrate that similar findings were generated when using different KDE bandwidths. The rank of data sources best correlated with the street survey geography were identical between output using 400m and 500m bandwidths. Similarly, there was very little difference when using a 300m bandwidth instead of 400m, although the Google BM geography outperformed that from Web extraction.

Altering the sensitivity of removal of data errors within the process also made little difference to the generated output. When fewer outliers were removed (minimum of 3 data points within 400m required), the rank of data sources most comparable with the street survey remained unchanged (Online Supplementary Table 6). Similarly, there was very little difference when the removal of outliers was increased (minimum of 10 data points within 400m), although the Flickr geography (the data source with the second highest level of agreement) outperformed that from Google BM (Online Supplementary Table 7). Under all scenarios of outlier removal, the highest correlation was found between the Street Survey and Web extraction geography.

Discussion

The discussion is structured around the two research questions presented in the introduction and our findings from the street survey and the generation of models using computational techniques and web-based sources of data.

- [RQ1]: To what extent do people agree on the 'membership' of landmarks within Sheffield City Centre and does the location at which a person is interviewed and their familiarity with an area affect their perception of the 'City Centre'?

From the results of a street survey and using the Fleiss Kappa statistic to measure agreement between peoples' judgements on the membership of landmarks ("inside", "outside", "on the boundary" and "don't know where it is") in Sheffield City Centre we find agreement varies from 'fair' to 'moderate'. Agreement measured using the Mann-Whitney U Test between locations shows a stronger correspondence. Overall, the landmark with the highest number of judgements of being inside Sheffield City Centre is the City Hall; whilst the Hallamshire Hospital is the landmark with the highest number of judgements of being outside the City Centre. We find the survey responses are only marginally influenced by the subject's current location. We also found that at the Railway Station there was more disagreement between the people interviewed at that location. We observe that the length of time someone has lived or worked in Sheffield has some influence on the results, with those living/working <5 years in Sheffield having consistently lower agreement than those living/working ≥ 5 years from all locations. Based on generating KDE representations of Sheffield City Centre from the responses of 61 pedestrians in a street survey, we find that the location of interview does not appear to have a significant effect on the resulting model. This latter result aligns with the lack of significant differences between the street survey responses at the different locations.

- [RQ2]: To what extent do the representations of 'City Centre' derived using data collected manually differ from or agree with those automatically gathered from web-based sources?

In the latter part of this paper we showed the results of representing Sheffield City Centre using an automated technique based on mining data from various web-based sources. Compared to existing studies we investigated multiple resources and the resulting spatial extents produced using KDE. The advantage of such an approach is speed: manual methods for deriving extents for vague regions are very time consuming and therefore less viable as a way to gather data for many areas. Another advantage of the approach is cost. Interrogating people on the street requires careful preparation and is both costly and labour intensive.

Various types of web-based sources have been utilised and the resulting models indicate inherent biases as shown by the differing geographies in Figure 2. The resulting representation based on mining data from Gumtree, which is based on local advertising data, is quite different from the other models. This reflects the bias in the Gumtree point locations towards more residential areas than the areas of leisure, shopping and tourism which were defined in the street survey. In contrast, the data based on Flickr.com appears slightly skewed eastwards and is more centred around the train station and surrounding area. It should be noted that the continuous KDE representation from the Gumtree data was most closely aligned with corresponding data from Rightmove and Google CM sources – which all, to some degree, reflect a more residential focus. Additionally, it should be noted that the street survey is heavily influenced by the presence or absence of suitable landmarks, as evident in Figure 2, in which the indentation on the north west of the representation could be attributed to the lack of a well-known landmark in that area.

A limitation of our approach is that our data do not relate to precisely the same time periods (Gumtree and Google data relate to 2007, whilst Flickr, Rightmove and web extracted data relate to 2016/2017). It should be stressed that whilst data might be extracted at a particular point in time they are likely to include historic data. For example, Flickr data are not really from a single temporal snapshot but relate to all data up to that point in time. Such issues are complex and require further work beyond the scope of this research to explore in more detail. Despite this, however, the overall agreement between the different data sources is encouraging (for example the highest level of agreement was found between Google BM and web extraction data despite being derived from different time periods – see Table 6). The work of Brindley (2006) demonstrated the overall general stability in Sheffield's neighbourhood definitions between 2012 and 2014 – with only 5% change in geography (recorded as the number of cells that were named differently between the two periods).

There was generally high agreement between some of the web-based representations of the city centre with the geographic extent obtained from our street survey and with the sources of crisp boundaries. This was reflected in (i) comparisons based on linear regression of cells from the continuous KDE surfaces, with 76% and 73% of the variance in the street survey output explained by the address-based web extraction and Google BM KDE surfaces respectively (see Table 6); and (ii) comparisons between crisp definitions derived from KDE decile thresholds of web-based data and existing crisp definitions, with high F_1-scores from Precision and Recall tests for both address-based web extraction and Google BM geographies against City Alert Scheme (85% and 79% respectively) and Lüscher & Weibel (2013) boundaries (89% and 79%).

This implies that automated approaches may indeed be suitable for deriving vague regions, as indicated by some of the previous studies. Our comparison of several different web-mined sources shows, however, that very different footprint definitions can be obtained from different sources, confirming that when using web-sourced data attention needs to be paid to the biases of the different sources. The clearest bias was found in the two sources that included data for property transactions (Gumtree and Rightmove) which had the lowest correlation with the street survey data. The results presented here also demonstrate that geographies based on the aggregation of different web-based sources may be beneficial in inferring region representations with a greater likelihood of agreement and consensus.

Conclusions and Future Work

This paper investigates approaches for modelling the extent of Sheffield City Centre in the UK. Subjects in a street survey were asked to judge the membership of 38 landmarks in central Sheffield. People were interviewed at three different locations in central Sheffield to establish to what extent people's perception of 'City Centre' was affected by location. We also assessed the effects of people's familiarity with Sheffield on their perception of the extent of the 'City Centre'. Finally, Kernel Density Estimate (KDE) models derived from subjects' responses in the street survey were compared to KDE models derived using automated approaches based on mining various web-based sources each of which generated sets of points associated with the City Centre.

Overall we observed a general agreement on the core of Sheffield City Centre based on interviewing subjects at three different locations in Sheffield. Our study has shown, however, that in this case location of the respondent seems to have little influence on people's definition of 'Sheffield City Centre'. Results based on using Fleiss' Kappa statistic to measure inter-annotator agreement did not find strong agreement about the membership of landmarks among respondents who were familiar with Sheffield City Centre (as defined by the length of time lived or worked in Sheffield). At the Railway Station there was a higher number of subjects unfamiliar with Sheffield compared to the two other locations. The automated method used to compute a representation of the City Centre boundary from web sources shows promising results, albeit varying with respect to the underlying data source. The best two approximations to the region obtained from our street survey data were found using address-based information extraction from the web and Google data on businesses contributed places, with the address-based data giving the best agreement (with an R-squared value of 0.76). The Google business data also had very close agreement with the web address-based information extraction methods perhaps explained by the fact that the Google data are also strongly oriented to structured addresses. However, the region obtained from the Gumtree web advertising data had very little agreement with the street survey data. This appeared to be due to the bias in the naming of city regions for advertising purposes such as rented accommodation. Notably lower agreement with street survey data was also found with data from the Rightmove real estate web site. This highlights that one must be careful when selecting web-based sources and be aware of potential biases in the sources. Our study shows that this may be at least partially offset by combining the geographies produced from a set of different web-derived data sources to identify the locations in common between the sources.

However, there are limitations in the work presented in this paper. One issue is that landmarks are mostly located inside and outside of the shopping area of Sheffield City Centre. The interview did not allow us to check the perception of the City Centre with respect to residential areas despite one respondent considering these areas to be a part of Sheffield City Centre. This is partly seen in the boundary obtained from Gumtree which contains information from a local advertising database and which differs considerably from other boundaries in extending beyond the others into more residential areas. We also gathered limited contextual factors to assess their impact on people's perception of Sheffield City Centre. However, people who go to Sheffield City Centre for shopping purposes are primed by defining 'City Centre' in terms of shopping; a person that is job hunting might shift their attention to the area of the 'City Centre' where the Job Centre, recruitment agencies or potential employer are located; a person looking for residential accommodation in central Sheffield might define the 'City Centre' in terms of residential areas (as reflected in the Gumtree data). Notably, the footprint of Sheffield City Centre derived from the street survey is more to the east than the web-based footprints. This may be due to a bias introduced by an area where suitable landmarks could not be found, although it may be that there is a difference in perception of the City Centre for the people in this survey.

Future work aims to gather more contextual information and assess the effects on people's perception of the City Centre. We also plan to experiment further with the web mining techniques, particularly with respect to detecting biases in the underlying data sources and assessing factors such as provenance, authoritativeness and trustworthiness. Approaches that use unstructured text and semi-structured data sources will also be tested, such as blogs, wikis and web pages. This may involve automatic extraction of geo-references, and assignment of spatial coordinates.

Acknowledgements

The authors like to thank Ordnance Survey (CASE/CAN/06/67) for funding this research on representation of place for geographic information retrieval. This work has also been partly funded by the EC FP6-IST 045335 TRIPOD project and partially supported by the EPSRC.

References

Arampatzis, A., van Kreveld, M., Reinbacher, I., Jones, C. B., Vaid, S., Clough, P., Joho, H., & Sanderson, M. (2006). Web based delineation of imprecise regions. *Computers, Environment and Urban Systems*, 30(4):436–459.

Bettencourt, L. M. (2013). The origins of scaling in cities, *science, American Association for the Advancement of Science*, 340:1438-1441.

Brindley, P. (2016) Generating vague geographic information through data mining of passive web data. Thesis (PhD). School of Computer Science, The University of Nottingham, UK.
Brindley, P., Goulding, J. & Wilson, M.L. (2017) Generating vague neighbourhoods through data mining of passive web data, *International Journal of Geographical Information Science*, *32(3):498-523.*
Brunsdon, C. (1995). Estimating probability surfaces for geographical point data: an adaptive kernel algorithm, *Computers & Geosciences*, 21(7):877–894.
Burrough, P. (1996). *Geographic Objects with Indeterminate Boundaries*, chapter Natural Objects with Indeterminate Boundaries, 3–28. Taylor and Francis.
Carletta, J. (1996). Assessing agreement on classification tasks: The kappa statistic, *Computational Linguistics*, 22(2):249–254.
Chen, J., & Shaw, S. L. (2016). Representing the Spatial Extent of Places Based on Flickr Photos with a Representativeness-Weighted Kernel Density Estimation. In J. A. Miller, D. O’Sullivan, & N. Wiegand (Eds.), Geographic Information Science: 9th International Conference, GIScience 2016, Montreal, QC, Canada, September 27-30, 2016, Proceedings (pp. 130-144).
Cioffi-Revilla C. (2017) Computation and Social Science. In: Introduction to Computational Social Science. Texts in Computer Science. Springer, Cham
Couclelis, H. (2003). The certainty of uncertainty: GIS and the limits of geographic knowledge, *Transactions in GIS*, 7(2):165–175.
Davies, C., Holt, I., Green, J., Harding, J., & Diamond, L. (2009). User needs and implications for modelling vague named places, *Spatial Cognition & Computation*, 9(3):174–194.
Evans, A. J. & Waters, T. (2007). Mapping vernacular geography: Web-based GIS tools for capturing 'fuzzy' or 'vague' entities, *International Journal of Technology, Policy and Management*, 7(2):134–150.
Fisher, P. (2000). Sorites paradox and vague geographies, *Fuzzy Sets Syst.*, 113(1):7–18.
Fisher, P. (2004). Where is Helvellyn? Fuzziness of multi-scale landscape morphometry, *Transactions of the Institute of British Geographers*, 29(1):106–128.
Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters, *Psychological Bulletin*, 76(5):378–382.
Frank, A. U. (1996). *Geographic Objects with Indeterminate Boundaries*, chapter The Prevalence of Objects with Sharp Boundaries in GIS, pages 29–40. Taylor and Francis.
Gao, S., Janowicz, K., Montello, D.R., Hu, Y., Yang, J-A., McKenzie, G. Ju Y., Gong, L., Adams, B. & Yan, B. (2017) A data-synthesis-driven method for detecting and extracting vague cognitive regions, *International Journal of Geographical Information Science* 31(6):1245-1271.
Glasersfeld, E. v. (1996). *Radikaler Konstruktivismus: Ideen, Ergebnisse, Probleme*. Frankfurt am Main: Suhrkamp.

Goodchild, M. F. (2000). GIS and transportation: Status and challenges, *Geoinformatica*, 4(2):127–139.
Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography, *GeoJournal*, 69(4):211–221.
Harpring, P. (1997). Proper words in proper places: the thesaurus of geographical names, *MDA Information*, 2(3):5-12.
Hill, L., Frew, J., & Zheng, Q. (1999). Geographic names. the implementation of a gazetteer in a georeferenced digital library, *D-Lib Magazine*, 5(1). Available online: http://www.dlib.org/dlib/january99/hill/01hill.html
Hollenstein, L. (2008). Capturing Vernacular Geography from Georeferenced Tags. MSc thesis: University of Zurich
Hollenstein, L. & Purves, R. (2010). Exploring place through user-generated content: Using Flickr tags to describe city cores, *Journal of Spatial Information Science*, 1(2010):21-48.
Jones, C. B., Purves, R. S., Clough, P. D., & Joho, H. (2008). Modelling vague places with knowledge from the web, *International Journal of Geographic Information Systems*, 22(10):1045–1065.
Landis, R. J. & Koch, G. G. (1977). The measurement of observer agreement for categorical data, *Biometrics*, 33(1):159–174.
Li, L. & Goodchild, M. F. (2012). Constructing places from spatial footprints. In *Proceedings of the 1st ACM SIGSPATIAL International Workshop on Crowd sourced and Volunteered Geographic Information - GEOCROWD '12*, pages 15–21.
Lüscher, P. & Weibel, R. (2013). Exploiting Empirical Knowledge for Automatic Delineation of City Centres from Large-scale Topographic Databases, *Computers, Environment and Urban Systems*, 37(1): 18-34.
Mackaness, W. A. & Chaudhry, O. Z. (2011). Automatic Classification of Retail Spaces from a Large Scale Topographic Database, *Transactions in GIS*, 15(3):291– 307.
Masucci, A. P.; Arcaute, E.; Hatna, E.; Stanilov, K. & Batty, M., On the problem of boundaries and scaling for urban street networks, *Journal of The Royal Society Interface, The Royal Society*, 12(111). Available online: http://rsif.royalsocietypublishing.org/content/12/111/20150763.long
Montello, D. R., Goodchild, M. F., Gottsegen, J., & Fohl, P. (2003). Where's downtown?: Behavioral methods for determining referents of vague spatial queries. *Spatial Cognition and Computation*, 3:185–204.
Montello, D., Friedman, A., & Phillips, D. (2014). Vague cognitive regions in geography and geographic information science, *International Journal of Geographical Information Science*, 28(9):1802–1820.
[O'Sullivan & Unwin, 2002] O'Sullivan, D. & Unwin, D. (2002). *Geographic Information Analysis*. Wiley.
Piaget, J. P. & Inhelder, B. (1948). *Die Entwicklung des räumlichen Denkens beim Kinde*, volume 6 of *Gesammelte Werke, Studienausgabe*. Klett-Cotta,

Stuttgart, 3. auflage, 1999 edition. translated from the French original "La reprÃ¨sentation de l'esapace chez l'enfants" by Heipcke, R.
Purves, R., Clough, P., & Joho, H. (2005). Identifying imprecise regions for geographic information retrieval using the web. In *GIS Research UK 13th Annual Conference*.
Purves, R.S., Clough, P., Jones, C.B., Arampatzis, A., Bucher, B., Finch, D., Fu, G., Joho, H., Syed, A.K., Vaid, S., & Yang, B. (2007). The design and implementation of SPIRIT: a spatially aware search engine for information retrieval on the Internet, *International Journal of Geographical Information Science*, 21(7):717–745.
Robinson AI, Carnes F, Oreskovic NM (2016). Spatial Analysis of Crime Incidence and Adolescent Physical Activity, *Preventive medicine*, 85:74-77.
Schockaert, S. & De Cock, M. (2007). Neighborhood restrictions in geographic IR. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '07*, page 167.
Schockaert, S., Cock, M. D., & Kerre, E. E. (2005). Automatic acquisition of fuzzy footprints. In *On the Move to Meaningful Internet Systems 2005: OTM 2005 Workshops, OTM Confederated International Workshops and Posters (SeBGIS 2005)*, volume 3762 of *LNCS*, Springer, 1077–1086.
Schockaert, S. (2011). Vague regions in geographic information retrieval. *SigSpatial Special*, 3(2), 24-28.
[Silverman, 1986] Silverman, B. (1986). *Density estimation: for statistics and data analysis*. Chapman and Hall, London.
Spelke, E. S. (1990). Principles of object perception, *Cognitive Science*, 14:29–56.
Tezuka, T. & Tanaka, K. (2005). Landmark extraction: A web mining approach, In Cohn, A. G. & D., M., editors, *Conference on Spatial Information Theory - COSIT 2005*, volume 3693 of *LNCS*, Springer, 379–396.
Thurstain-Goodwin, M. & Unwin, D. (2000). Defining and delineating the central areas of towns for statistical monitoring using continuous surface representations, *Transactions in GIS*, 4(4):305–317.
Twaroch, F. A., Jones, C. B., & Abdelmoty, A. I. (2008). Acquisition of a vernacular gazetteer from web source, In *LOCWEB '08: Proceedings of the first international workshop on Location and the web*, 61–64, New York, NY, USA. ACM.
Twaroch, F.A., Purves, R.S., & Jones, C.B. (2009). Stability of Qualitative Spatial Relations between Vernacular Regions Mined from Web Data. In *Proceedings of Workshop on Geographic Information on the Internet*, Toulouse, France, April 6th, 2009.
Vögele, T., Schlieder, C., & Visser, U. (2003). Intuitive modelling of place name regions for spatial information retrieval, In Kuhn, W., Worboys, M. F., & Timpf, S., editors, *Spatial information theory: foundations of geographic information science International Conference, COSIT 2003, Kartause Ittingen,*

Switzerland, Sept. 24-28, volume 2825 of *Lecture Notes in Computer Science*, 239–252, Berlin. Springer Verlag.

Location	The Moor		Peace Gardens		Railway Station	
Respondents	Male	Female	Male	Female	Male	Female
	5	17	9	9	11	10
Total	22		18		21	

Table 1: Location and numbers of respondents

	All responses (%)	The Moor (*N*=22)	Peace Gardens (*N*=18)	Railway Station (*N*=21)
Mappin Street	29 (47.5%)	10 (45.5%)	9 (50.0%)	10 (47.6%)
Campo Lane	27 (44.3%)	7 (31.8%)	6 (33.3%)	14 (66.7%)
Lady's Bridge	23 (37.7%)	5 (22.7%)	6 (33.3%)	12 (57.1%)
Shalesmoor	23 (37.7%)	5 (22.7%)	7 (38.9%)	11 (52.4%)
Kelham Island	22 (36.1%)	9 (40.9%)	4 (22.2%)	9 (42.9%)
Devonshire Green	20 (32.8%)	8 (36.4%)	5 (27.8%)	7 (33.3%)
Fire/Police Museum	20 (32.8%)	7 (31.8%)	5 (27.8%)	8 (38.1%)
Fire Service HQ	20 (32.8%)	5 (22.7%)	5 (27.8%)	10 (47.6%)
Park Hill	20 (32.8%)	3 (13.6%)	5 (27.8%)	12 (57.1%)

Table 2: Most unknown ('don't know') places in central Sheffield across all responses and by interview location

Landmarks judged 'Inside'	
	Responses (%)
City Hall	59 (98.3%)
Crucible	58 (96.7%)
Town Hall	58 (96.7%)
Cathedral	56 (93.3%)
Winter Garden	56 (93.3%)
Fargate	54 (90.0%)
Moor	52 (86.7%)
Castle Market	51 (85.0%)
Bus Station	49 (81.7%)
Fitzalan Square	46 (76.7%)

Landmarks judged 'Outside'	
	Responses (%)
Hallamshire Hospital	50 (83.3%)
Sheffield University	40 (66.7%)
Ice Rink	40 (66.7%)
Weston Park	38 (63.3%)
Sheffield United FC	37 (61.7%)
Wicker	29 (48.3%)
Shalesmoor	29 (48.3%)
Waitrose	27 (45.0%)
Park Hill	26 (43.3%)
Kelham Island	25 (41.7%)

Table 3: 'Inside' and 'outside' judgments across all three locations

Kappa (κ)	$\kappa<0$	$0\leq\kappa<0.2$	$0.2\leq\kappa<0.4$	$0.4\leq\kappa<0.6$	$0.6\leq\kappa<0.8$	$0.8\leq\kappa<1$
Interpretation (agreement)	Poor	Slight	Fair	Moderate	Substantial	Almost Perfect

Table 4: Interpretation of the Kappa score (Landis & Koch, 1977)

		The Moor	Peace Gardens	Railway Station	All
	Subjects	22	18	21	61
	κ	0.357 (fair)	0.502 (moderate)	0.246 (fair)	0.077 (slight)
≥5 years in Sheffield	Subjects	12	11	10	33
	κ	0.450 (moderate)	0.589 (moderate)	0.389 (fair)	0.485 (moderate)
<5 years in Sheffield	Subjects	10	7	11	28
	κ	0.314 (fair)	0.316 (fair)	0.223 (fair)	0.066 (slight)

Table 5: Agreement between participants' decisions and locations

	Flickr	Google BM	Google CM	Gumtree	Rightmove	Web extraction	Street Survey
Flickr	-	36.4	48.8	1.3	12.8	51.6	70.3
Google BM		-	74.3	9.7	73.7	85.1	72.9
Google CM			-	18.6	52.4	74.6	66.2
Gumtree				-	20.7	9.7	1.4
Rightmove					-	67.1	48.3
Web extraction						-	75.9
Street Survey							-

Table 6: Comparison of similarity of the KDE surfaces from the R-squared values derived using linear regression

Decile contours:	Flickr	Google BM	Google CM	Gumtree	Rightmove	Web extraction	Street Survey
10%	67.8	48.6	49.7	**36.8**	58.0	61.0	50.7
20%	72.3	54.2	54.5	35.5	63.0	67.5	56.2
30%	**73.5**	59.7	59.4	35.2	67.4	73.2	63.0
40%	73.3	65.1	63.1	35.2	69.7	78.2	70.8
50%	71.6	72.0	65.8	35.8	**70.8**	82.5	80.1
60%	67.9	76.7	68.1	35.6	69.1	**85.0**	**86.3**
70%	60.5	**78.8**	**70.8**	33.9	65.1	80.9	84.6
80%	48.9	70.3	65.3	29.4	55.3	64.2	70.1
90%	27.4	47.3	43.7	15.4	36.6	38.1	44.3

Table 7: F_1-score (from Precision and Recall): decile contours for each data source against City Alert Scheme crisp definition

Decile contours:	Flickr	Google BM	Google CM	Gumtree	Rightmove	Web extraction	Street Survey
10%	**73.3**	67.6	64.2	**49.4**	79.7	82.8	71.0
20%	70.7	73.0	64.9	47.0	83.7	87.2	75.1
30%	67.9	75.8	**65.6**	44.6	86.5	**88.8**	79.5
40%	64.8	77.6	65.3	41.5	**87.0**	87.3	84.3
50%	60.2	**78.9**	63.3	39.7	82.9	81.4	**86.8**
60%	53.5	75.3	62.3	36.9	73.7	72.5	80.1
70%	44.6	68.9	60.7	33.2	61.9	59.3	67.6
80%	32.2	54.3	50.0	27.7	45.9	43.8	50.1
90%	17.2	31.8	28.5	16.3	25.7	24.5	28.9

Table 8: F_1-score (from Precision and Recall): decile contours for each data source against Lüscher & Weibel (2013) crisp definition

Number of computational sources in agreement	Definition from City Alert Scheme	Definition from Lüscher & Weibel (2013)	Definition from 50% contour from Street Survey
1+	47.1 (30.8, 100.0)	30.9 (18.3, 100.0)	41.9 (26.5, 100.0)
2+	59.9 (42.8, 100.0)	40.6 (25.5, 100.0)	53.9 (36.9, 100.0)
3+	74.9 (60.0, 99.7)	52.7 (35.8, 100.0)	68.3 (51.8, 100.0)
4+	84.6 (74.4, 98.1)	62.1 (45.1, 100.0)	77.1 (63.7, 97.5)
5+	**86.1 (89.3, 83.2)**	**72.5 (59.4, 93.1)**	**79.2 (76.1, 82.4)**
6	70.0 (96.6, 54.9)	61.2 (62.6, 59.8)	62.4 (78.5, 51.7)

Table 9: F_1-score (Precision and Recall in brackets) comparison between two crisp definitions and differing level of agreement with the aggregation of computational sources

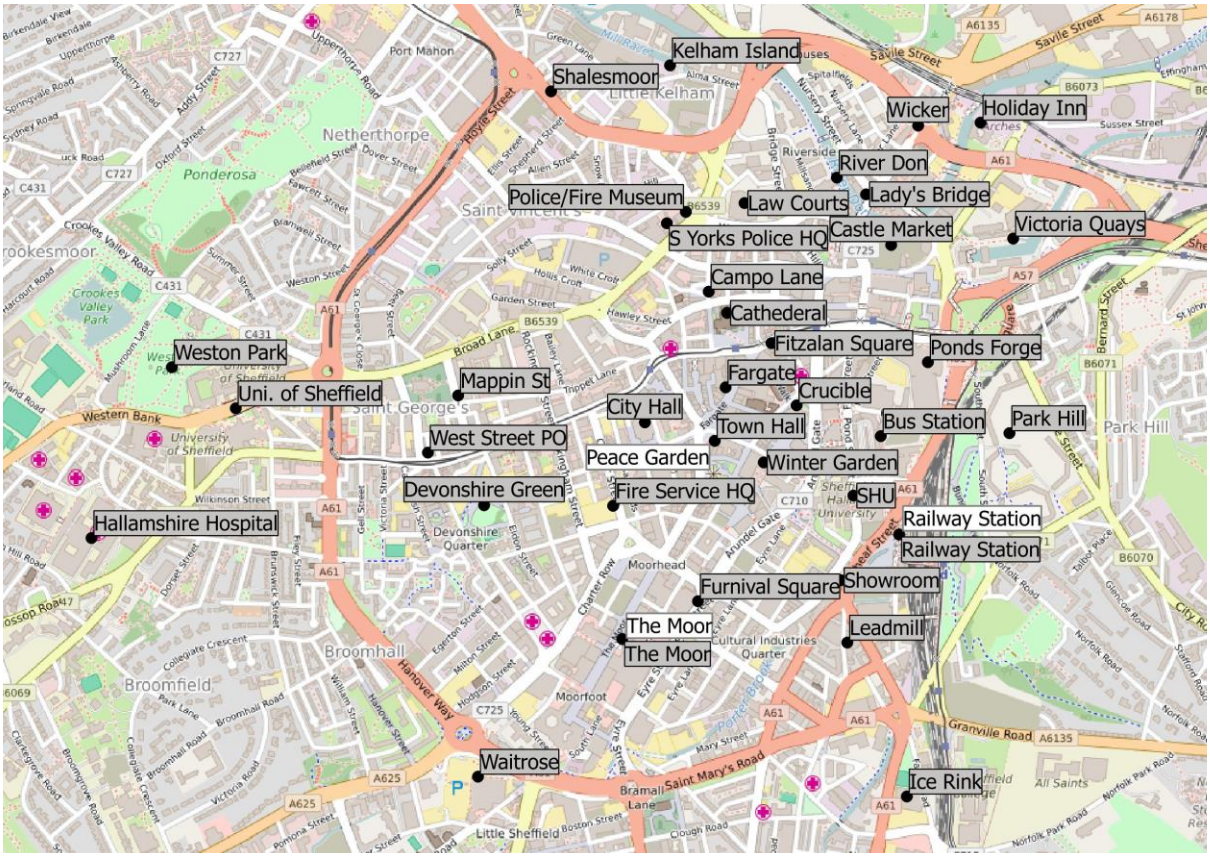


Figure 1: Landmark (grey boxes) and interview (white boxes) locations

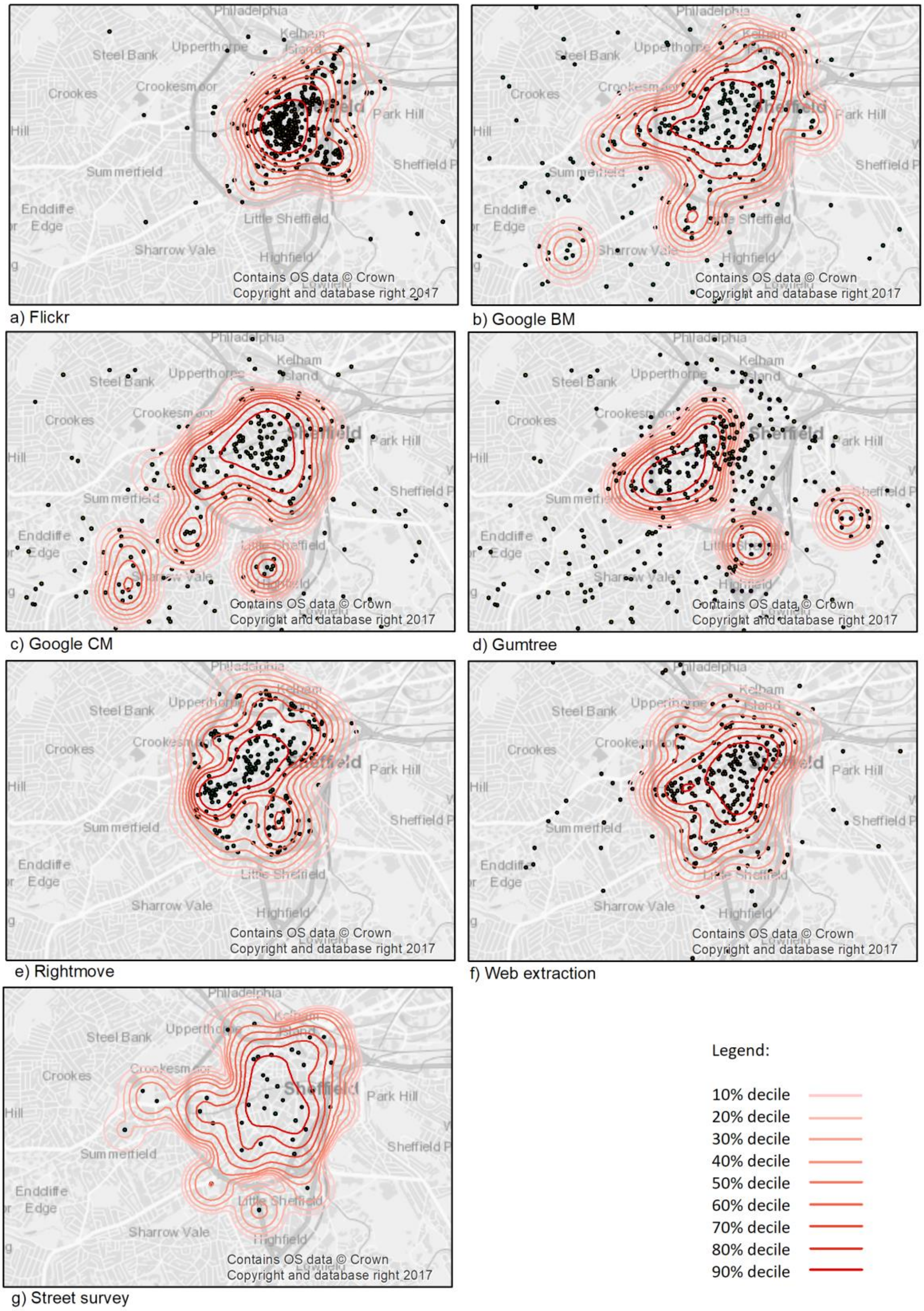


Figure 2: KDE output as decile contours for the various data sources

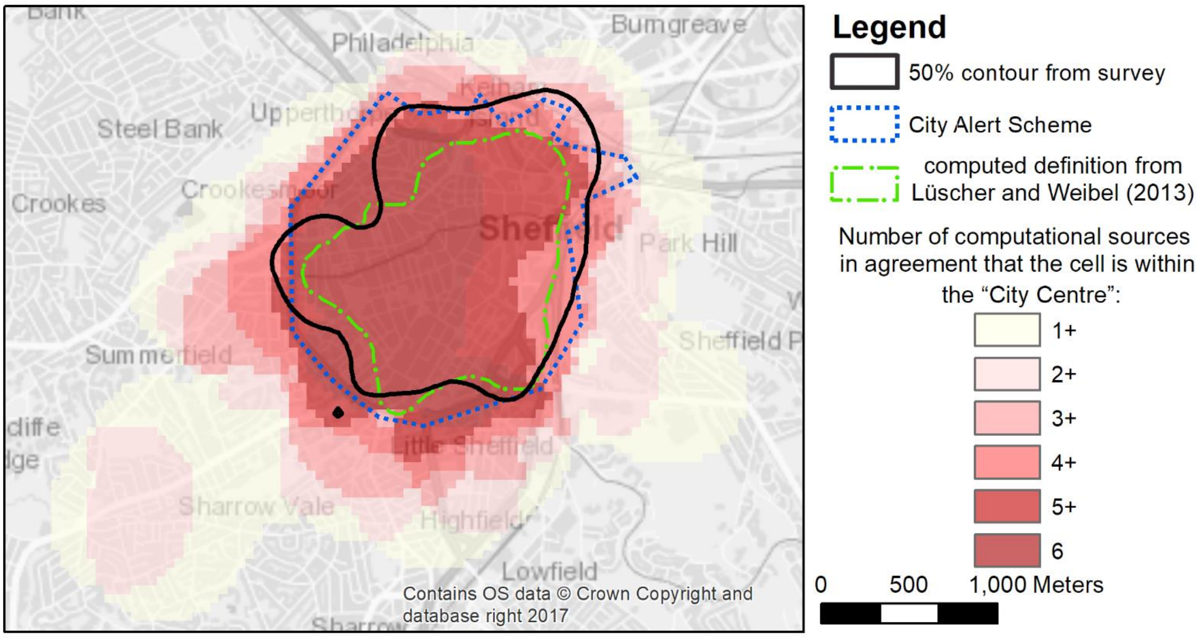


Figure 3: Aggregation of the KDE surfaces from the different computational sources

Supplementary online only material

Decile contours:	Flickr	Google BM	Google CM	Gumtree	Rightmove	Web extraction	Street Survey
10%	**73.3** (70.9, 76.0)	67.6 (51.6, 97.7)	64.2 (50.6, 87.7)	**49.4** (48.9, 49.8)	79.7 (67.6, 97.1)	82.8 (72.0, 97.4)	71.0 (55.8, 97.6)
20%	70.7 (75.6, 66.3)	73.0 (59.4, 94.6)	64.9 (54.4, 80.3)	47.0 (50.4, 44.0)	83.7 (75.4, 94.2)	87.2 (81.0, 94.5)	75.1 (62.3, 94.7)
30%	67.9 (80.4, 58.8)	75.8 (65.5, 90.0)	**65.6** (59.0, 73.8)	44.6 (52.4, 38.8)	86.5 (83.0, 90.2)	**88.8** (88.6, 89.0)	79.5 (70.7, 90.9)
40%	64.8 (85.4, 52.2)	77.6 (72.0, 84.3)	65.3 (63.3, 67.3)	41.5 (54.7, 33.5)	**87.0** (89.9, 84.4)	87.3 (94.8, 80.9)	84.3 (81.4, 87.4)
50%	60.2 (90.8, 45.0)	**78.9** (80.4, 77.4)	63.3 (68.4, 58.9)	39.7 (59.4, 29.8)	82.9 (95.0, 73.5)	81.4 (98.1, 69.6)	**86.8** (93.8, 80.8)
60%	53.5 (95.3, 37.2)	75.3 (87.3, 66.2)	62.3 (76.5, 52.5)	36.9 (63.7, 26.0)	73.7 (97.1, 59.4)	72.5 (100.0, 56.9)	80.1 (98.6, 67.5)
70%	44.6 (99.7, 28.7)	68.9 (94.7, 54.2)	60.7 (90.7, 45.6)	33.2 (71.0, 21.7)	61.9 (99.4, 45.0)	59.3 (100.0, 42.1)	67.6 (99.8, 51.1)
80%	32.2 (100.0, 19.2)	54.3 (99.0, 37.4)	50.0 (100.0, 33.3)	27.7 (83.4, 16.6)	45.9 (100.0, 29.8)	43.8 (100.0, 28.1)	50.1 (100.0, 33.4)
90%	17.2 (100.0, 9.4)	31.8 (100.0, 18.9)	28.5 (100.0, 16.6)	16.3 (90.2, 8.9)	25.7 (100.0, 14.8)	24.5 (100.0, 14.0)	28.9 (100.0, 16.9)

Online Supplementary Table 1: F1-score (with Precision and Recall in brackets): decile contours for each data source against City Alert Scheme crisp definition

Decile contours:	Flickr	Google BM	Google CM	Gumtree	Rightmove	Web extraction	Street Survey
10%	67.8 (52.7, 95.1)	48.6 (32.1, 100.0)	49.7 (33.4, 97.2)	**36.8** (29.2, 50.0)	58.0 (41.0, 99.2)	61.0 (43.9, 100.0)	50.7 (34.0, 100.0)
20%	72.3 (60.7, 89.5)	54.2 (37.1, 100.0)	54.5 (38.2, 94.9)	35.5 (29.8, 43.8)	63.0 (46.5, 97.7)	67.5 (51.0, 100.0)	56.2 (39.1, 100.0)
30%	73.5 (66.6, 82.0)	59.7 (42.7, 99.3)	59.4 (43.8, 92.2)	35.2 (31.7, 39.5)	67.4 (52.1, 95.3)	73.2 (58.3, 98.5)	63.0 (46.1, 99.7)
40%	**73.3** (72.2, 74.3)	65.1 (49.0, 97.2)	63.1 (49.2, 87.9)	35.2 (34.7, 35.8)	69.7 (56.9, 89.9)	78.2 (66.3, 95.3)	70.8 (55.0, 99.3)
50%	71.6 (78.7, 65.7)	72.0 (58.0, 94.8)	65.8 (55.6, 80.6)	35.8 (39.1, 33.0)	**70.8** (62.6, 81.5)	82.5 (75.8, 90.5)	80.1 (67.6, 98.0)
60%	67.9 (85.6, 56.2)	76.7 (68.1, 87.9)	68.1 (63.5, 73.4)	35.6 (43.7, 30.1)	69.1 (68.1, 70.1)	**85.0** (86.9, 83.2)	**86.3** (80.6, 92.8)
70%	60.5 (92.6, 44.9)	**78.8** (79.8, 77.9)	**70.8** (77.2, 65.4)	33.9 (50.0, 25.7)	65.1 (75.3, 57.4)	80.9 (97.5, 69.1)	84.6 (91.5, 78.8)
80%	48.9 (100.0, 32.4)	70.3 (89.2, 58.0)	65.3 (91.0, 51.0)	29.4 (58.5, 19.6)	55.3 (82.7, 41.5)	64.2 (100.0, 47.2)	70.1 (97.4, 54.7)
90%	27.4 (100.0, 15.8)	47.3 (97.9, 31.2)	43.7 (100.0, 27.9)	15.4 (53.9, 9.0)	36.6 (92.1, 22.9)	38.1 (100.0, 23.5)	44.3 (100.0, 28.4)

Online Supplementary Table 2: F1-score (with Precision and Recall in brackets): decile contours for each data source against Lüscher & Weibel (2013) crisp definition

Decile contours:	Flickr	Google BM	Google CM	Gumtree	Rightmove	Web extraction
10%	74.2 (66.9,83.3)	63.7 (46.7,100.0)	58.6 (43.8,88.2)	**43.4 (40.0,47.4)**	73.0 (58.4,97.4)	76.1 (62.3,97.9)
20%	**74.5 (73.9,75.2)**	69.8 (53.8,99.4)	62.0 (49.1,84.1)	42.6 (42.3,42.8)	76.9 (65.0,94.3)	80.3 (69.8,94.5)
30%	73.9 (80.4,68.3)	73.9 (60.1,95.9)	64.6 (54.5,79.1)	41.2 (44.5,38.3)	79.1 (70.9,89.4)	**82.1 (76.2,89.0)**
40%	72.2 (87.0,61.8)	78.0 (67.7,92.0)	**65.7 (59.5,73.4)**	39.5 (47.5,33.8)	**79.9 (76.6,83.5)**	81.0 (81.3,80.6)
50%	65.4 (89.4,51.5)	**82.1 (77.8,86.9)**	64.9 (64.9,64.9)	38.5 (52.2,30.4)	79.0 (83.4,75.0)	79.0 (87.4,72.0)
60%	57.3 (91.8,41.6)	80.7 (86.2,75.9)	64.5 (72.7,57.9)	36.7 (57.0,27.1)	73.0 (87.9,62.5)	74.4 (93.5,61.8)
70%	48.5 (96.6,32.4)	75.9 (95.1,63.1)	63.9 (86.7,50.6)	33.3 (63.7,22.5)	64.6 (93.8,49.3)	65.6 (99.8,48.8)
80%	36.5 (100.0,22.3)	60.7 (99.5,43.6)	55.0 (98.5,38.1)	28.2 (75.1,17.4)	50.8 (98.7,34.2)	49.1 (100.0,32.6)
90%	19.7 (100.0,10.9)	36.0 (100.0,22.0)	32.3 (100.0,19.3)	16.0 (77.5,8.9)	29.3 (100.0,17.1)	27.9 (100.0,16.2)

Online Supplementary Table 3: F1-score (with Precision and Recall in brackets): decile contours for each of the computational data sources against a crisp definition based on the 50% contour from the Street Survey output

	Flickr	Google BM	Google CM	Gumtree	Rightmove	Web extraction	Street Survey
Flickr	-	30.5	41.2	3.3	6.0	42.9	55.8
Google BM		-	67.4	12.6	65.8	80.5	65.8
Google CM			-	11.6	41.0	64.6	54.9
Gumtree				-	14.9	4.3	0.0
Rightmove					-	56.5	31.4
Web extraction						-	60.7
Street Survey							-

Online Supplementary Table 4: Sensitivity analysis - Comparison of similarity of the KDE surfaces from the R-squared values derived using linear regression using a KDE bandwidth of 300m

	Flickr	Google BM	Google CM	Gumtree	Rightmove	Web extraction	Street Survey
Flickr	-	45.3	54.4	0.2	20.4	57.7	78.3
Google BM		-	0	22.9	81.8	90.6	79.6
Google CM			-	25.0	60.3	81.1	72.2
Gumtree				-	25.6	15.4	3.8
Rightmove					-	74.6	59.0
Web extraction						-	83.5
Street Survey							-

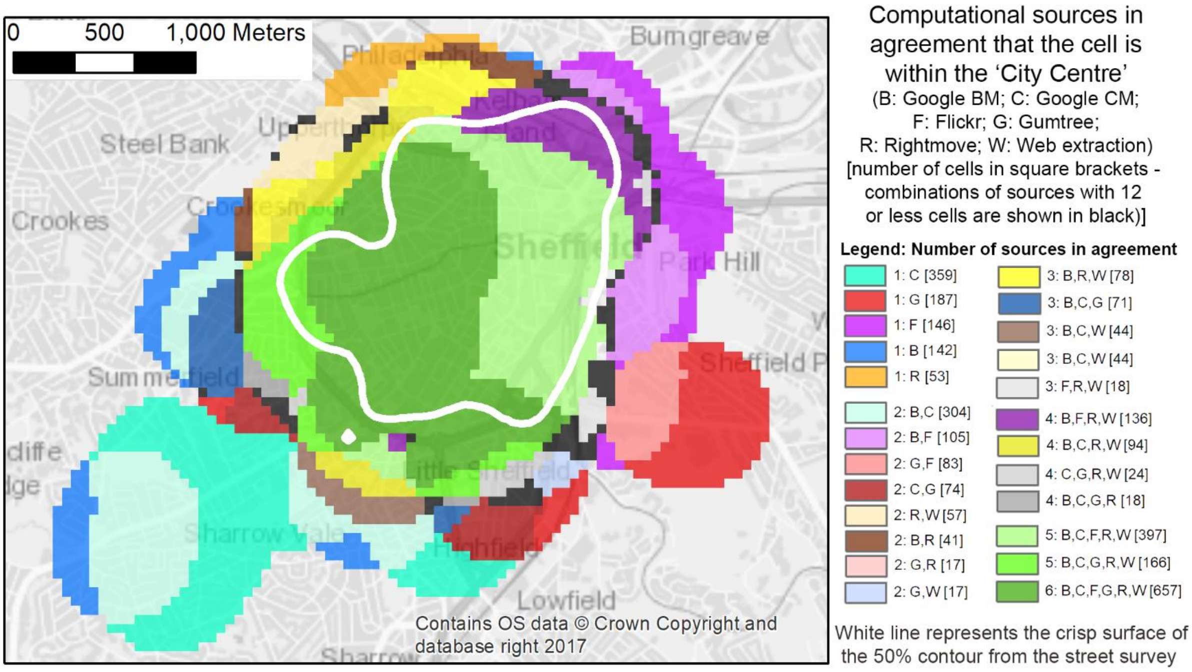
Online Supplementary Table 5: Sensitivity analysis - Comparison of similarity of the KDE surfaces from the R-squared values derived using linear regression using a KDE bandwidth of 500m

	Flickr	Google BM	Google CM	Gumtree	Rightmove	Web extraction	Street Survey
Flickr	-	39.1	50.4	0.0	17.9	55.6	71.6
Google BM		-	74.8	24.7	75.7	83.8	72.1
Google CM			-	22.1	58.9	73.3	68.7
Gumtree				-	58.9	17.1	3.6
Rightmove					-	71.9	29.5
Web extraction						-	77.1
Street Survey							-

Online Supplementary Table 6: Sensitivity analysis - Comparison of similarity of the KDE surfaces from the R-squared values derived using linear regression using a lower level of outlier removal (3 points within 400m)

	Flickr	Google BM	Google CM	Gumtree	Rightmove	Web extraction	Street Survey
Flickr	-	29.7	40.6	n/a	12.2	54.1	69.8
Google BM		-	85.2	n/a	72.3	80.9	68.9
Google CM			-	n/a	55.9	77.9	64.0
Gumtree				-	n/a	n/a	n/a
Rightmove					-	57.8	49.5
Web extraction						-	76.8
Street Survey							-

Online Supplementary Table 7: Sensitivity analysis - Comparison of similarity of the KDE surfaces from the R-squared values derived using linear regression using a higher level of outlier removal (10 points within 400m)



Online Supplementary Figure 1: Aggregation of the KDE surfaces from the different computational sources