

Methods for Neighbourhood mapping, boundary agreement

Abstract

Any analytical study of a neighbourhood must begin with an accurate definition of the geographic region that contains it. For a long time, there has been an interest in taking surveys of neighbourhood extents, but this can generate numerous haphazardly sketched polygons. Researchers typically face the challenges of using boundary polygons reported by each participant and unifying these polygons into one representative boundary. Over the years, several researchers have reported their findings on methods for unifying these boundaries. We present and compare the following five methods (two existing, one modified and two new): Dalton radial average, Bae–Montello average, a vectorised version of the Bae–Montello raster grid overlay, a vectorised derivative inspired by the Wenhao kernel density axis method (MKDA) and a new k-medians clustering method.

A crowdsourced evaluation method is presented. N=42 raters ranked the five methods of aggregating real boundary data based on the results from three study areas. We found that the boundary aggregation method derived from the Bae–Montello grid, closely followed by the Dalton radial average method, provided the most reasonable results. This paper outlines the reasons for these results and illustrates how this knowledge may point to the ability of future algorithms to improve the presented methods. The paper ends with a recommendation that neighbourhood boundaries should utilise boundaries derived from the Bae–Montello raster grid overlay method and/or the Dalton radial average method to facilitate comparisons in the field.

Introduction

Ahlbrant (Ahlbrandt, 1984) suggests that city dwellers strongly identify with their neighbourhoods. Neighbourhoods are something that influences their experience of the city and how they interact with its people. The planner and economist Webster (Webster, 2003) reported that the neighbourhood is a substantial factor for people when selecting new housing. The subjective experience of place by residents is referred to by geographers, social scientists, and humanities scholars as a “sense of place”. It refers to the experience of connection between people and their immediate environment (Tuan, 1974; Jorgensen and Stedman, 2001; Cresswell, 2004), leading to terms such as “place attachment” (Lewicka, 2011). Neighbourhoods have been associated with a range of positive psychological effects (Scannell and Gifford, 2017). At its core, the study of neighbourhood mixes our subjective and social experiences and memories with geographically specific places. Brown, Raymond and Corcoran (Brown et al., 2015) suggested that place attachment needs to be “mappable”, that is, it would benefit from a more detailed understanding of its spatial dimension.

Interest in identifying neighbourhood boundaries also exists for very practical reasons. For example, to aid a quick response from emergency services, operators must be able to translate from verbal neighbourhood references to known geographic locations: an injured person may be reported on the edge of the local “China Town”. In that case, this region needs to be given some explicit bounds to allow emergency services to attend quickly. While it would be tempting to use the administrative

unit, it has been often shown that administrative units significantly under- or over-estimate *self-defined neighbourhoods* (see Colabianchi et al., 2014; Coulton et al., 2013; Jenks and Dempsey, 2007).

The problem that emerges for the researcher is how to take several hand-drawn, self-reported neighbourhood boundaries and create a single representative polygon from them. Several prior attempts have been made to create such representative polygons (Bae and Montello, 2018; Dalton, 2007; Yan et al., 2000). To date, we can find no prior work that has attempted to both describe and evaluate these methods. In this paper, we aim to identify prior techniques, suggest refinements and provide comparisons between them.

Background

Past research on neighbourhoods can be retraced to Lynch's (Lynch, 1971) explorations, beginning with the collecting of subjective impressions of the extents of a certain Boston neighbourhood. Lynch was one of the first to begin asking residents to draw the extent of their neighbourhood on a map, along with other questions. Lynch reported that these regions never precisely aligned. The lack of agreement gives substance to the question of whether a neighbourhood has a spatial component, as suggested by Talen (Talen, 2000). However, it is often challenging to interpret such fuzzy data concretely.

In the 1970s, a psychologist (Lee, 1973) asked 219 housewives to "*please draw a line around the part which you consider as your neighbourhood or district*". He observed a good deal of variation, observing that eight polygons from residents 100 yards apart "*showed almost no coincidence*".

Following Lynch and Lee, Taylor (Taylor et al., 1984) and Guest (Guest and Lee, 1984) used interview methods based on approaching residents in a region. While slower, this method ensures that those interviewed are local and not interlopers passing through the neighbourhood from elsewhere. Guest reported that 25% of those interviewed thought that neighbourhood was a social or non-spatial concept. The remaining 75% believed it had some spatial aspect even if this was not its primary definition. In these cases, interpreting the regions reported was also challenging. This lack of certainty raises the question of whether these regions share some underlying consistency. Is there some shared notion of where a neighbourhood is or do the variations suggest that researchers are asking a question with no real answer?

To further unpick this, (Montello et al., 2003) extended previous work by asking participants to draw two boundaries. The first was the "certain-within-bounds" - the region that the participant was certain was within the downtown region. The second was the "likely-within-bounds", which identified the region the participant thought was *likely* to be downtown. Finally, the participants were asked to identify a point they felt was the centre of the downtown region. From this, it was possible to construct a dot-density map that represented the fuzzy nature of the impression of the downtown region.

In 2006 Clapp and Wang (Clapp and Wang, 2006) used a method based on sales transactions and their street addresses which were then aggregated through Classification and Regression Trees to create neighbourhoods based on observed behaviour, rather than a participant reflective response with a sketch map. While this avoids the issue of dealing with the overlapping areas, the results were never tested against residents' perceived neighbourhoods.

In 2007, Dalton (Dalton, 2007, 2011) introduced a new postal method for the collection of data concerning neighbourhood boundaries. This work also sought to convert several, collected sketch boundaries into a single representative one that could be used to compare reported boundaries to

proposed spatial correlations with the neighbourhood. Dalton introduced a simple geometric algorithm that reduced the boundaries to a single aggregate one. It did this by casting a ray from a common centre and using the average distance along each ray from the centroid to determine an average point. By rotating the ray around the common centroid, a polygon can be generated. Dalton also observed the regularity of boundaries, such as alignment to significant high streets, which suggests that neighbourhoods are geographically aligned rather than that they begin centred on the respondent's dwelling.

Montello's work could be seen as an inspiration for the 2015 paper by Yu et al. (2015), who proposed a method to find a central business district using kernel density estimation (KDE). Here, the methodological stages of Montello's work are converted to a KDE. The observation is that points in the urban fabric can be assigned to the central business district (a specific neighbourhood just like Montello's "downtown"). Yu, Ai, and Shaown's work further refined the KDE field by including distance along the road network (network KDE) rather than metric distance. The results were convincing. However, the evaluation of Yu, Ai, and Shaown's work consisted only of a comparison with a survey by Yan et al. (Yan et al., 2000).

In 2016, Campbell, Henly, Elliott, and Irwin (Campbell et al., 2009) constructed neighbourhood boundaries for four neighbourhoods in a western city. They did not attempt to create an aggregate boundary; they reported on a subjective neighbourhood definition that needed 50% of residents to agree that an area was within the neighbourhood boundary for it to be accepted as part of that neighbourhood. They observed that sometimes boundaries are identified along thoroughfares, thus dividing potentially more prominent neighbourhoods. Like Campbell, in 2016 Van Gent, Boterman, and Grondelle (Van Gent et al., 2016) used maps and significant landmarks to create a density diagram showing regions for 70%, 80% and 90% agreement. However, the individual boundaries were more beneficial for unpicking the social relationships relating to class and mutual inclusion than the larger boundaries.

In 2018, Bae and Montello (Bae and Montello, 2018) extended Dalton and Montello's work and, to an extent, Yu, Ai, and Shaown's KDE approach by proposing two new methods. After collecting sketch maps from 50 respondents living around LA's Koreatown neighbourhood, they created two new boundary aggregation methods. The first of these was a variant of the Dalton method. Here the centre of each boundary is found (unlike the Dalton method, which requires a single common centroid). While the rays do not align, they can be used to find two-dimensional points that can be similarly aggregated to find a new boundary.

The second Bae-Montello method can be seen as a high-density version of the Campbell method combined with the Yu, Ai, and Shaown KDE method. It relies on a raster grid overlaid on the geographic map. The grid represents a 10x10 ft (3x3 m) cell grid encompassing all the underlying sketch boundaries. A simple test is performed for each boundary to decide if the cell centre is inside or outside a particular boundary. If the cell centre is inside, then a counter for the cell is incremented. This is repeated for all cells and all input boundaries. By dividing the final count by the number of input boundaries, it is possible to end up with a percentage or probability that a particular cell is within the neighbourhood. This probability density field can then be converted into a single aggregate boundary by applying a filter (>75% agreement or >50% agreement) that keeps values greater than a particular percentage of agreement. Bae and Montello compared their findings to an administrative boundary from the Los Angeles City Council (LACC) and a crowd-sourced version from the MappingLA project and found a reasonable correlation between the two.

In 2019, psychologists Stülpnagel, Brand, and Seemann (Stülpnagel et al., 2019) performed a similar boundary collection process. This time they used a tool on the internet to collect the source boundaries. While they did not attempt to find a single aggregate, they did make two observations.

First, they conclusively demonstrated that the boundaries were not circles and that the resident's home was not at the centre of their boundary. This result suggests that there is some approximate, underlying, common region. If we take the hypothesis that a person's neighbourhood is literally where their neighbours are, we would expect that the majority of the polygon boundaries would be centred on the respondent and be approximately circular, given that there would be no particular bias for a neighbour in one direction over another. The fact that this was not found adds to the evidence that a common but intangible "neighbourhood" exists.

Research question

It can be seen from the literature that there have been several attempts to try to identify a single region. In each of the cases, the researcher in question has developed an independent boundary aggregation method. Up to this point, we can find no references to papers where an attempt has been made to compare previous methods. This paper aims to analyse a variety of techniques presented in the past literature and introduce new methods further developing them. Our objective is not to identify a single, 'correct' method but to give the reader a comparison from which an informed decision may be made. A further contribution of this paper is that a method is presented to objectively evaluate the various averaging methods. In doing so, we hope to identify the best methods to be used by those investigating neighbourhoods and inspire and inform future researchers wishing to develop similar measures. We hope to help identify methods that could be standardised to allow comparative analyses of different, reported neighbourhoods.

Neighbourhood algorithms compared

In this section, we will present five methods that were implemented for comparison. This section will detail these methods.

Dalton average radial

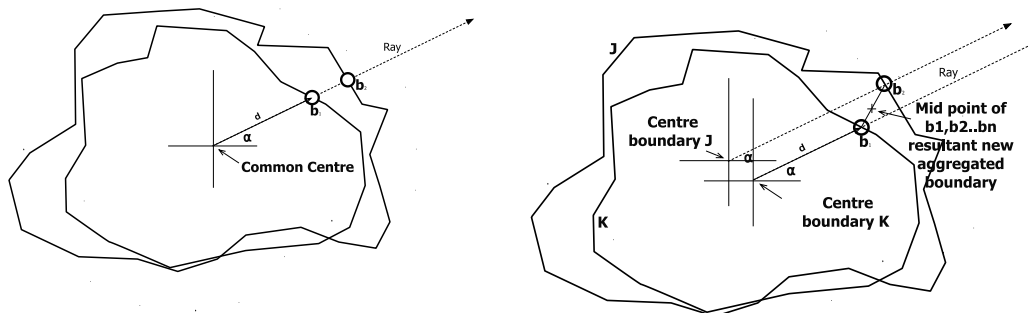


Figure 1a Common centre for Dalton Average 1b Bae distinct centre method

This method is an implementation of the approach employed by Dalton's (Dalton, 2007, 2011) average area method, in which an operator selects a common centre for all input boundaries. A ray at angle α is cast from this centre; the method exploits the fact that boundaries tend to be convex and only intersect the ray once. The distance along the ray α of the intersection point with boundary i creates a distance d_{ci} along the ray from the starting point. This is repeated for all input boundaries, and the ray and the mean of distance \bar{d} are found. The distance is then projected back on to the ray to find a point in space \bar{p}_α (see Supp Fig 1). The angle from the starting point of the ray is then angle α , and the process is repeated, forming a path that traces a new aggregate

boundary. See Figures 2a and S1 for an example. One criticism of this method was that the operator selected a common centre. As can be seen in the Appendix figure S6, changes of the common centre lead only to small changes in the boundary.

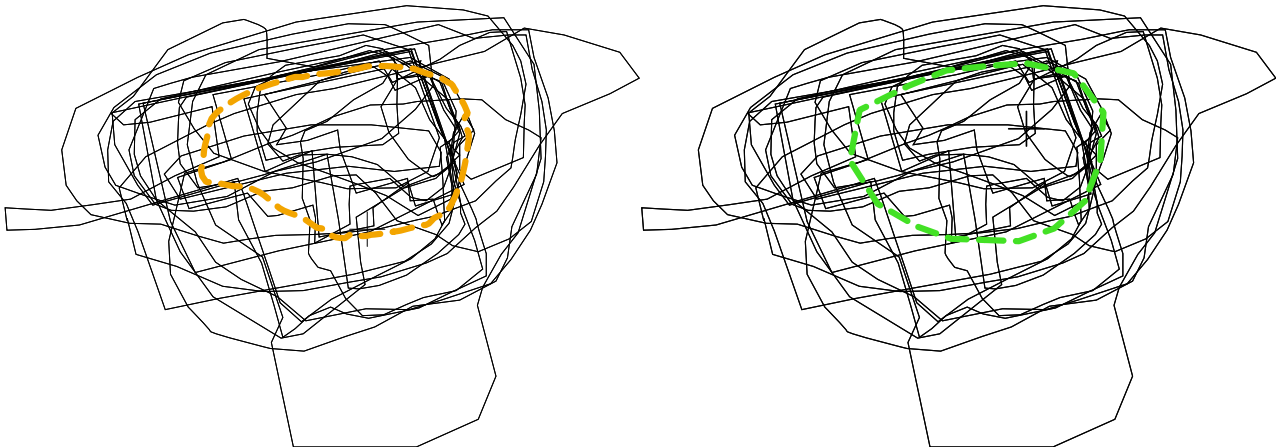


Figure 2a Dalton Average for Brentham Garden Suburb(left). 2b Bae Average boundary for Brentham Garden Suburb(right)

Bae and Montello's average radial intersect method (Bae-Montello average)

As mentioned, the Bae and Montello radial intersect method (Bae and Montello, 2018) is an extension of the previous method that removes the arbitrary use of an operator selection of centre point. In this method, the centre of each boundary is used (see Figure 1b). This method has the advantage of precise repeatability over the Dalton method which accepts the slight differences introduced by the operator.

Their method allows for a region that possesses no common centre. The average radial intersect method begins with the centroid (mean centre) of each boundary and projects a ray out for a given angle α . This ray intersects with the boundary as in the Dalton method, and this intersection point P_α gives a typically two-dimensional point in space. This intersection process is repeated for each boundary, resulting in a point field. The coordinates for the point field are averaged to find the average point in space. As with the Dalton method, the angle α is incremented and the process repeated, tracing the path, which is the aggregate boundary. In the original work, only 16 rays were used compared to the 90 used in the Dalton method. To improve comparative clarity with the other methods, 180, 2°-angle rays were used. See Figure 2b, figure S2 for an example.

Modified Bae and Montello raster grid overlay method (Bae-Montello grid)

This method is a variant of the Bae-Montello raster grid overlay method (Bae and Montello, 2018). It was created to provide results that would be compatible with the other methods presented, explicitly giving a single boundary. The method begins as per the Dalton average radial method by tracing a ray from a common centre. The boundary intersects with the ray, creating a distance from the ray origin dd_{ai} . In this case, rather than being average, the values are stored in ascending order. For an aggregate boundary at the 75% agreement level, the distance of the 75% item in order is found. For 100 items, the 75th smallest value is chosen; for 50 items, the 37th smallest distance d_{ai} is selected. This method will always pick an actual value for one of the boundaries rather than lying between two other boundaries. This method relies on selecting the level of agreement (>50% or >75% in Bae and Montello) before processing. Once a point has been found along the ray, the angle α is incremented and the process repeated, tracing a path, which is the aggregate boundary. This variant was chosen to allow a direct comparison with other methods. The original method was also implemented, but we could not observe any examples in our dataset of the two methods producing different results. The principal advantage of this method over the original raster grid is that it produces a discrete boundary that can be used for later analysis. As mentioned, it does require the use of a common centre point, which the original method does not. For clarity, this method will be

referred to as the ‘modified grid method’ or a ‘modified raster method’ even though no raster information was used in its creation. See Figures 3 and S3 for an example.

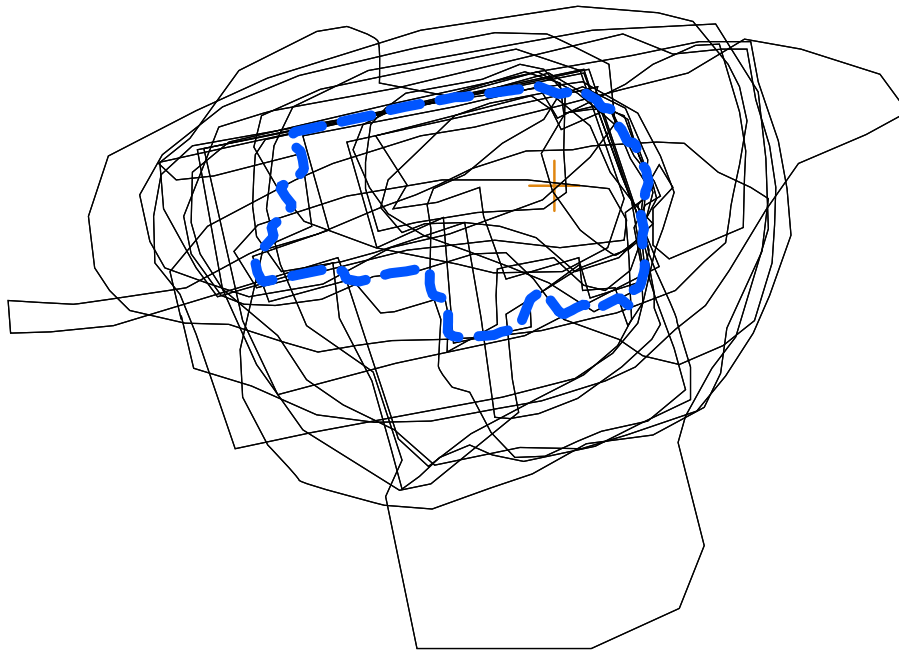


Figure 3 Sample boundary Modified Bae-Montello raster grid method for Brentham Garden Suburb

New methods

This paper will now introduce two original methods developed specifically for this paper. The intention was to create alternative mechanisms to aggregate the boundaries together, to allow a comparison to be performed.

Radial K-Median

This method is based on methods found in artificial intelligence and machine learning. It begins with the Dalton method by selecting user defined centre and finding a distance for each boundary along a ray. These values can be summarised in several ways. In this case, each point of intersection is also a point along a line. In the grid method, it is assumed that when sufficient people agree that a point is within the space, a point with 50% or 75% agreement identifies the boundary. In this case, the region about which most parties agree is the boundary, is chosen as the boundary. If many respondents agree that a road is the edge of the neighbourhood, the presumption is that this forms part of the boundary. As with the raster method, it is presumed that a zone of agreement suggests that the aggregate boundary is most likely present. This reduces the effects of outliers at both extremes. In this method, the distances along the ray are entered into a histogram of n equal-sized bins. The bin with the highest number of values is used; this is converted into $\overline{d_\alpha}$ by finding the average of all values in the bin. It is possible to remove the effect of bin size by recomputing this point with multiple bin sizes. In our experimentation, a value of $n = 9$ was found to produce similar results. As with the Dalton method, the $\overline{d_\alpha}$ along the path is converted back to a point in space, and a path is traced to create several bounds. See Figures 4a and S4 for an example.

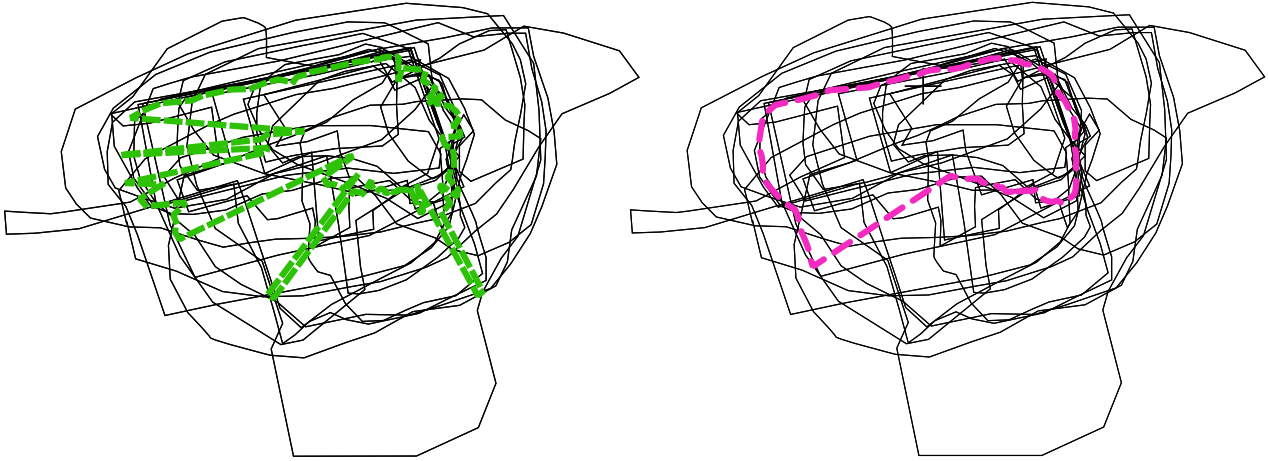


Figure 4a Radial K-Median boundary (left), 4b Maximum Kernel Density Axis Measure boundary(right) for Brentham Garden Suburb

The maximum kernel density axis (MKDA)

This new method was inspired by Wenhao Yu, Tinghua Ai, and Shiwei Shao (Yu et al., 2015). Yu's method was originally designed for a central business district region from point-based observations, with the CPD being another neighbourhood but not a residential one. The Yu method used an underlying street grid rather than a Cartesian distance, which is unavailable in this case. No claims were given in the original paper about the validity of this method as a mechanism for finding neighbourhood boundaries.

In this method, we begin with the Dalton process of selecting a centroid and casting a ray from that centroid. Each intersection of the ray with sketch boundaries $I_{\overline{d}_{\alpha}}$ is regarded as a Gaussian point of maximal likelihood. That is, the drawn boundary is viewed as a sample of the "true" underlying boundary. Using a linear kernel density function, it is possible to find the point that is maximally likely to be that point. The kernel density estimator is expressed as:

$$f(s) = \sum_{i=1}^n \frac{1}{h} K\left(\frac{s - c_i}{h}\right)$$

where $f(s)$ is the estimated density value at location s , n is the number of points (boundaries), h is the search bandwidth, and $s - c_i$ is the distance between the event point c_i and the location. K is a weight function called the kernel function; in our case, we used a Gaussian function, where:

$$K(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-u)^2}{2\sigma^2}}$$

The intersection point is treated as a likelihood function with a decaying value. There is a maximum value of likelihood at the intersection point, but this drops off as the distance from the point is selected. A variance parameter controls the drop-off. From this point of view, all intersections contribute to finding the final value.

By considering this process as one of maximising likelihood, we can also introduce other factors that might influence the final boundary. In this case, by using probability, we can also introduce the distance of the previous radius of the previous ray $\overline{d}_{\alpha-1}$ using a constant, C , to control how much influence the previous value has. The influence of the previous value would smooth any discontinuity caused by the source-boundaries crossing. As before, the angle of the ray is then swept round to generate the final boundary.

This method has the same disadvantages as the Dalton average ray method in that it needs at least one point of universal commonality between the input boundaries as the other methods do. As with all methods, 180 rays were used with a separation of 2° between each ray. See Figures 4b and S5 for an example.

Evaluation

There is no ‘gold standard’ against which the averaged boundaries can be compared (Campbell et al., 2009). For example, government boundaries or mapping boundaries are primarily imposed on neighbourhoods for administrative purposes. A government agency typically ensures that neighbourhoods align with adjacent neighbourhoods: for administrative purposes, there can be no gaps between neighbourhoods nor overlaps. Nevertheless, there is no evidence to support the notion that, for the lived experience of a district, neighbourhoods necessarily abut to their surrounding neighbourhoods. Two adjacent neighbourhoods may be permeable, with one overlapping another.

In previous research, an aggregate neighbourhood boundary has been a stepping stone for further analysis. Given the essentially subjective nature of the neighbourhood, the general approach has been to experiment with several different methods and choose one that seems to most accurately represent the data presented (Bae and Montello, 2018; Dalton, 2011). As several researchers have mentioned (Bae and Montello, 2018; Dalton, 2007; Yu et al., 2015), it is clear that specific methods appear to produce polygons that more strongly match our expectations of where the neighbourhood truly lies. Given the subjective nature of this, the solution this paper presents is to use a comparative, lay-person evaluation of several different algorithms to select the most effective (i.e. that most matches people’s intuition).

Our evaluation process was as follows. Each of the five methods was implemented in the same software and used to process each of three different neighbourhood datasets. The dataset was derived from the data originally used by Dalton (Dalton 2011). Data was collected using the postal map method originally described by Dalton (Dalton, 2011). Here, approximately 400 A5 cards are posted through the letterboxes of potential participants in the target neighbourhood. The card contains a short introductory text, a sketch map and two additional questions asking for the name of the neighbourhood and the number of years the participants had lived there. Participants are asked, “*On this map, would you please draw a line around the area you think of as your neighbourhood? By neighbourhood, I simply mean the part of the city where you live*” (Dalton 2011). They are asked to draw an X to represent the approximate position of their street and then return the card by Freepost.

Three datasets were surveyed in London in the UK: Hampstead Garden Suburb, Clerkenwell and Brentham Garden Suburb: Hampstead Garden Suburb (Zone H) which elicited 34 boundaries and Brentham Garden Suburb (Zone B) which resulted in 23 boundaries in this dataset are two well-established suburban neighbourhoods and Clerkenwell (Zone C), with 27 boundaries, is an inner city neighbourhood. Each dataset consisted of the set of self-reported sketch boundaries for one neighbourhood. The same data was used by each method, resulting in its own representative neighbourhood boundary.

For each neighbourhood dataset, the five methods were presented anonymously in pairs to 42 non-expert but college-educated raters. The raters were shown both the single average network boundary outlines and the original input boundaries. For each of the possible pairings, the raters provided a preference. We hypothesised that raters would prefer those that they subjectively felt were more representative of the average neighbourhood boundary. If there was a clear preference for one boundary method over others, it would suggest that the method had managed to embody some aspect that supported their intuition. By using several alternatives, not explicitly explaining the algorithm and by using independent evaluators, it was felt that such a system could reduce bias.

The null hypothesis, in this case, is that it is not possible to have one method that is more representative than others. In which case the methods would produce no preferences beyond what would be expected by random choice.

Before starting the raters were given the following instructions., “*The mechanism of this evaluation is largely subjective and based on your expert (but intuitive) judgment. For each question, you will be presented with two alternative algorithmically computed boundaries. These two boundaries are shown in a colour, along with the original neighbourhood boundaries (from the residents, and used as input into the algorithm) shown as black-line polygons. All you have to do is select which coloured boundary you think is the better summary of ALL of the black-line polygons.*

We are asking that your judgement be made on a scale of 1-5, which is intended to measure your level of certainty. A value of 1 or 5 means you are confident that one coloured boundary is a better representation (of the black polygons) than the other, competing, coloured boundary. A value of 3 suggests that you judge that neither boundary is better than the other (they are both equally good or equally poor). Values of 2 and 4 suggest that one boundary is marginally superior to the other.”

In our experiment, there were three separate regions. The pair of scoring-images were presented grouped by the region. Before presenting the pairs, a single image was shown (see Figure 5)



Figure 5 Example of reported boundaries, each colour represents a different reported boundary with underlying street map in black..

This map represents each sketch boundary, drawn in a separate colour to allow it to be distinguished from the others. It also included a simplified street map representing the underlying street layout, see Figure 5. To avoid bias, the origins and locations of the neighbourhood in question were withheld from the raters. After showing the street map and individual boundaries, a series of pairwise boundary comparisons was presented to the raters.

The core method presented human raters with different pairs of neighbourhood boundaries in a survey. By presenting each pair, overlaid on the original input boundaries, as above, it would be possible to judge which was intuitively more representative of the input boundaries, see Figure 6 for a sample.

Which of these thick, coloured polygons best represents the full set of thinner, black lines? If one representation is clearly better than the other, select 1 or 5, if one representation is only slightly better, select 2 or 4 and if they are equally as good then select the middle value, 3. *

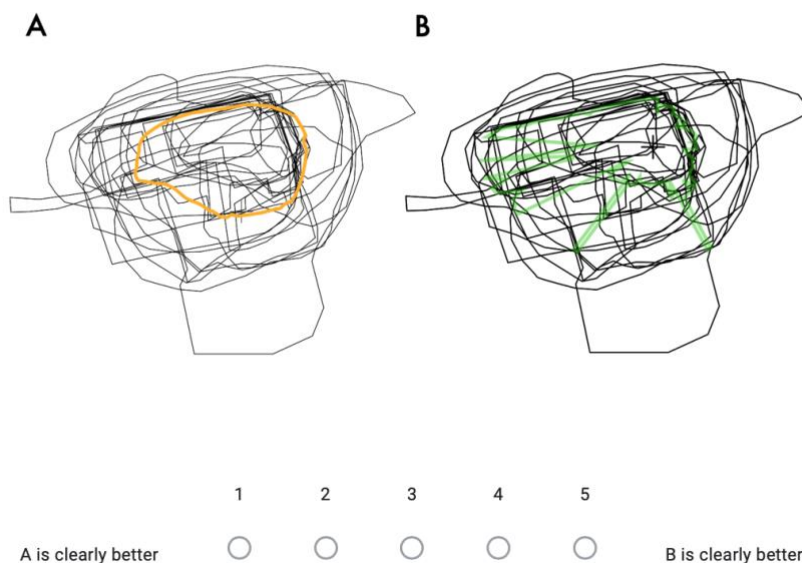


Figure 6 Example of the side-by-side comparison of two possible algorithms presented to the human raters in the experiment. A and B present average and K-Medians average (green right), respectively.

Below each pair of maps, a Likert scale with five possible settings was presented. The scale offered the options “A is clearly better”, “A is slightly better”, “neutral”, “B is slightly better” and “B is clearly better”. The use of the neutral value allowed raters to express no preference if neither method appeared to capture the original inputs adequately. To avoid presentation-sequence bias, each of the pairings was presented randomly. Each algorithm was given its own colour and letter and, again to avoid bias, the algorithm itself was not mentioned.

For the pairwise choice, each algorithmically-produced polygon was visualised, on its own map, as a thick coloured line superimposed on the many, thin, black lines representing the original individual sketch boundaries (see Figure 6). Data was presented online using survey software, allowing the study to reach several raters and not be limited by geography. After presenting all pairwise choices for each neighbourhood, a space was left for the raters to leave comments or questions. In the data reported in this paper, five methods and three different datasets were evaluated this way.

For each neighbourhood dataset, the five methods were exhaustively put together as pairs. This gave 25 possible pairs. Five of the possible pairings could be removed because it would require the method to be compared with itself. Of the 20 remaining methods, half could be eliminated due to duplicate comparisons. For example, if method A was compared to method B, it was not necessary to compare method B with method A. This left ten possible pairwise comparisons per dataset. For the three datasets, this gave rise to 30 comparisons, not overly fatiguing the raters, while still giving a variety of neighbourhoods to be compared.

Study

After seeking and obtaining ethical permission from the host university, the survey was hosted on Google Forms, and a website was used to obtain 42 lay raters. Raters were filtered by requiring them to have at least a graduate level of education. This was done to ensure that those taking part could read a map and not be forced to choose randomly. Additionally, raters were also filtered by requiring them to have high scores from previous unrelated experiments. This was to ensure that

those taking part would engage with the experiment. All of the results were examined, and any that showed a consistent answer pattern, for example the first option selected for all answers, would have been excluded. In this case, it turned out not to be necessary.

As a test for one of the neighbourhoods, one comparison was repeated. As mentioned in the methodology, the questions were presented in random order. The duplicate question would be easy to overlook, thus testing if the respondents were consistent in their choices. If respondents answered the repeat comparison significantly differently, then the rater could be eliminated from the study. No limitation on geographic location for raters was used, allowing a random spread of raters globally, reducing confounds due to cultural backgrounds.

Analytic approach

The first hypothesis tested whether the human raters consistently noticed any difference between the methods. The null hypothesis being that the human raters could not identify any difference between the presented systems, and the scores were no different from random choices. For the statistical analysis, any preference for one method over another (“1” or “2” on the Likert scale, indicating a preference for one, versus “4” or “5” as a preference for the other) was noted as a vote. Any neutral scores of “3” were discarded. Votes for each method could be tallied against all method pairs presented for all three test maps. This would give a relative frequency of votes for each method. These frequencies could then be tested against those expected for a purely random set of choices using a chi-squared test.

The second method was to look for differences between methods by using the votes on a per participant level. To consider the individual scores, each participant’s preferences were calculated as the number of times each participant voted for that specific algorithm. So for P3 this was: Dalton average = 4; K-Median = 3; Bae-Montello average = 0; Grid method = 10; Kernel density = 4. Indicating that P3 preferred the method Dalton average over others four times but the Grid method ten times. This summary provided a column of 42 values per method which could then be compared statistically for differences between the values.

Results

To examine the first hypothesis: if the human raters were consistently noticing any difference between the methods, table 1 shows a tally of votes for each neighbourhood and for all neighbourhoods in total. Using a chi-squared test (see table 2), it is possible to compare the results for both individual and overall cases compared with the expected random values, if evenly distributed between the algorithms. It can be seen from table 2 that in all cases both r- the p-values are below the threshold of 0.01 showing that the changes in frequency cannot be attributed to random variance suggesting the values express some consistency between all participants.

Method	Brentham GS	Clerkenwell	Hampstead GS	All
Dalton average	83	113	54	250
Bae-Montello average	50	98	51	199
Bae-Montello Grid	120	90	101	311
Radial K-Medians	35	21	39	95
MKDA	74	58	89	221

Table 1 Binary votes preferences used in the chi-squared test

Method	Brentham GS	Clerkenwell	Hampstead GS	All
Chi-squared value	53.137	71.03	42.64	116.79
Critical value $p=0.01$	13.27	13.27	13.27	13.27
p-value	<0.001	<0.001	<0.001	<0.001

Table 2. Chi-squared test results

To answer the question of whether there is any significant difference between the votes for each method, the data was coded to give votes on a ‘per participant’ level. This produced five columns (one per method) with each with 42 votes. Performing a Shapiro-Wilk test using Jamovi v1.6 (The jamovi project, 2021) (Team, 2020) it could be observed that four methods had a $p > 0.01$ except for K-Median ($W=0.790$ p value < 0.01). This requires the use of a non-parametric paired Wilcoxon signed-rank test. Performing these, all the tests have a significance $p < 0.05$ suggesting that all the values were significantly different from each other. See the appendix Table S1 for all values.

The median values were: 6 for Dalton average radial; 1.5 for K-Median; 3.5 for Bae-Montello Average; 8.0 for Grid Method and 4.5 for the Kernel Density method. From this it can be seen that the K-Medians process is the worst-performing of all the methods. The MKDA method, while sophisticated, overall takes third place. Surprisingly, the Dalton average and the Bae-Montello average produce significantly different results, and overall, the Dalton average is lower and the highest median is elicited by the modified Bae-Montello raster grid method. Looking in more detail at the results for individual regions, it is possible to see that the raster grid method is very successful in both the Brentham GS and Hampstead GS areas. In the case of Clerkenwell, this method falls into third place but still performs very well.

Discussion

In this section, this paper will examine the methods in light of the evaluation process. It seems significant that the K-Medians algorithm was widely perceived as being inferior to the others. For future algorithmic development, it would appear that having unstable values that emerge as spikes in the final polygon is something to be avoided. The MKDA method does appear to pick areas of strong consensus but the areas of weak agreement perhaps results in a polygon that makes it appear unrepresentative. It can be seen that, in the case of Clerkenwell, the Bae-Montello raster grid-based method shows some minor instabilities to the northeast of the polygon. This may have contributed to its underperformance in this particular situation. The Dalton average does moderately well and appears in bifurcated systems, such as Clerkenwell, to also perform relatively well indicating its robustness.

Overall, it cannot be denied that the method derived from the Bae-Montello (Bae and Montello, 2018) raster grid stands out as people’s preferred method. This reinforces the work of Bae and Montello (Bae and Montello, 2018), who expressed their preference for this method in an American gridded test case. This has now been reinforced with a European non-gridded and historic set of neighbourhoods. The question is, what is it about this algorithm that appeared to match the intuition of these raters? Due to its design, the algorithm will always trace one of the existing drawn boundaries rather than being a composite of different boundaries. For regions with substantial streets acting as boundaries, such as Hampstead Garden Suburb, the algorithm managed to pick up where many people had drawn a relatively firm edge. It also did not appear to be overly affected by significant outliers. Where it did badly was where it resulted in an occasional spike or rapid inflexion, which appeared to make the results less natural looking.

Conclusions

The first contribution of this paper is the use of human raters to compare methods in an unbiased and reproducible way. The statistical tests have shown that naive human raters give a consistent preference for one algorithm over another. We found, in our experiment, that all of the human judges gave genuine preferences, certainly suggesting their diligence. We also found that our human judges were relatively consistent. That is, while they might not make exactly the same judgement in the duplicate case, they never contradicted their previous decisions. We were conscientious in ensuring that all methods appeared essentially different.

The prime contribution to knowledge of this paper was to bring together several algorithms and conduct a comparative analysis between them. In general, all the algorithms, except possibly the K-Medians clustering algorithm, performed well enough for general use. It can be seen that the original Dalton average method works relatively well. However, the modified Bae-Montello raster grid does appear to result in neighbourhood-boundaries that are preferred by our human raters. This also has the advantage of being the only method which has been tested on an American city and so has the certainty of more generality. These two methods are recommended for researchers in the field as the primary analysis methods. It is also suggested that those developing new consensus methods should attempt to evaluate against these algorithms to allow a reasonable comparison. While the Bae-Montello raster grid does appear to give the best results, this is not meant to suggest that the algorithm is currently the best in all possible situations. In the discussion, we have suggested why we believe the algorithm appears to outperform others in the view of our subjective lay raters. It is certainly possible that further improvements could be made on these algorithms.

Limitations

It should be acknowledged that there are a number of limitations to this work. The first is that the regions selected for the comparison were all from one single city, namely London. This could introduce a bias towards one method, which may work well for European cities but not for others that have fundamentally different morphological types, such as cities typically found in America and in many rapidly developing areas. As mentioned in the conclusions, the Bae-Montello raster grid derived method was initially developed for an American setting.

Future work

As mentioned in the limitations, this work is based on comparisons of neighbourhoods collected from one city in Europe. Future work should attempt to reproduce the process of collecting neighbourhood sketch boundaries from several different cities worldwide. It would be particularly helpful to create an open dataset to be used by researchers in the field to allow the cross-analysis of methods, particularly if such a dataset encompassed several neighbourhoods from different cultures around the world. There is also the assumption that the boundaries in this dataset all refer to a single neighbourhood and it is possible that several neighbourhoods (overlapping, adjacent or nested) could be present. Future work in this area will include a process to disaggregate such data into their separate neighbourhoods.

References

Ahlbrandt RS (1984) A Community Focus. In: *Neighborhoods, People, and Community*. Springer, pp. 1–11.

Bae CJ-H and Montello DR (2018) Representations of an Urban Ethnic Neighbourhood: Residents' Cognitive Boundaries of Koreatown, Los Angeles. *Built Environment* 44(2). Alexandrine Press: 218–240.

Brown G, Raymond CM and Corcoran J (2015) Mapping and measuring place attachment. *Applied Geography* 57. Elsevier: 42–53.

- Campbell E, Henly JR, Elliott DS, et al. (2009) Subjective constructions of neighborhood boundaries: lessons from a qualitative study of four neighborhoods. *Journal of Urban Affairs* 31(4). Taylor & Francis: 461–490.
- Clapp JM and Wang Y (2006) Defining neighborhood boundaries: Are census tracts obsolete? *Journal of Urban Economics* 59(2): 259–284. DOI: <https://doi.org/10.1016/j.jue.2005.10.003>.
- Colabianchi N, Coulton CJ, Hibbert JD, et al. (2014) Adolescent self-defined neighborhoods and activity spaces: Spatial overlap and relations to physical activity and obesity. *Health & place* 27. Elsevier: 22–29.
- Coulton CJ, Jennings MZ and Chan T (2013) How big is my neighborhood? Individual and contextual effects on perceptions of neighborhood scale. *American journal of community psychology* 51(1). Springer: 140–150.
- Cresswell T (2004) *Place: A Short Introduction*. Blackwell Publishing.
- Dalton N (2007) Is Neighbourhood Measurable? In: *6th International Space Syntax Symposium*, Istanbul, Turkey, 12 June 2007.
- Dalton N (2011) *Synergy, intelligibility and revelation in neighbourhood places*. University College London (University of London).
- Guest AM and Lee BA (1984) How urbanites define their neighborhoods. *Population & Environment* 7(1): 32–56.
- Jenks M and Dempsey N (2007) Defining the neighbourhood: Challenges for empirical research. *Town Planning Review* 78(2): 153–178.
- Jorgensen BS and Stedman RC (2001) Sense of place as an attitude: Lakeshore owners attitudes toward their properties. *Journal of environmental psychology* 21(3). Elsevier: 233–248.
- Lee T (1973) Psychology and living space. *Image and Environment: Cognitive Mapping and Spatial Behavior*: 87–108.
- Lewicka M (2011) Place attachment: How far have we come in the last 40 years? *Journal of environmental psychology* 31(3). Elsevier: 207–230.
- Lynch K (1971) *1960. The Image of the City*. Cambridge, MA: MIT Press.
- Montello DR, Goodchild MF, Gottsegen J, et al. (2003) Where's downtown?: Behavioral methods for determining referents of vague spatial queries. *Spatial Cognition & Computation* 3(2): 185–204.
- Scannell L and Gifford R (2017) The experienced psychological benefits of place attachment. *Journal of Environmental Psychology* 51. Elsevier: 256–269.
- Stülpnagel R von, Brand D and Seemann A-K (2019) Your neighbourhood is not a circle, and you are not its centre. *Journal of Environmental Psychology* 66: 101349. DOI: <https://doi.org/10.1016/j.jenvp.2019.101349>.
- Talen E (2000) The problem with community in planning. *Journal of Planning Literature* 15(2): 171.
- Taylor RB, Gottfredson SD and Brower S (1984) Neighborhood naming as an index of attachment to place. *Population & Environment* 7(2): 103–125.
- Team RC (2020) R: A language and environment for statistical computing [Internet]. R Foundation for Statistical Computing; 2018. Available from: [_ http://www.R-project.org/_](http://www.R-project.org/). [Google Scholar].
- The jamovi project (2021) *Jamovi (Version 1.6) [Computer Software]*. Available at: <https://www.jamovi.org>.
- Tuan YF (1974) *Topophilia: A Study of Environmental Perception, Attitudes, and Values*. Prentice-Hall.

- Van Gent WP, Boterman WR and van Grondelle MW (2016) Surveying the fault lines in social tectonics; Neighbourhood boundaries in a socially-mixed renewal area. *Housing, Theory and Society* 33(3). Taylor & Francis: 247–267.
- Webster C (2003) The nature of the neighbourhood. *Urban Studies* 40(13): 2591–2612.
- Yan X, Zhou C and Leng Y (2000) Functional features and spatial structure of CBDs in Guangzhou. *ACTA GEOGRAPHICA SINICA-CHINESE EDITION*- 55(4). SCIENCE PRESS: 485–493.
- Yu W, Ai T and Shao S (2015) The analysis and delimitation of Central Business District using network kernel density estimation. *Journal of Transport Geography* 45. Elsevier: 32–47.