RESEARCH ARTICLE

μ-shapes: Delineating urban neighborhoods using volunteered geographic information

Matt Aadland, Christopher Farah, and Kevin Magee

InnoVision Basic and Applied Research Office, National Geospatial-Intelligence Agency, USA

Abstract: Urban neighborhoods are a unique form of geography in that their boundaries rely on a social definition rather than a well-defined physical or administrative boundary. Currently, geographic gazetteers capture little more than then the centroid of a neighborhood, limiting potential applications of the data. In this paper, we present μ-shapes, an algorithm that employs fuzzy-set theory to model neighborhood boundaries suitable for populating gazetteers using volunteered geographic information (VGI). The algorithm is evaluated using a reference dataset and VGI from the Map Kibera Project. A confusion matrix comparison between the reference dataset and μ-shape's output demonstrated high sensitivity and accuracy. Analysis of variance indicated that the algorithm was able to distinguish between boundary and interior blocks. This suggests that, given the existing state of GIS technology, the μ-shapes algorithm can enable neighborhood-related queries that incorporate spatial uncertainty, e.g., find all restaurants within the core of a neighborhood.

Keywords: boundary delineation, urban neighborhoods, volunteered geographic information, spatial footprint, vague spatial region

1 Introduction

Operationally, neighborhood boundaries can be defined by the degree of consensus between individuals. Ethnographic studies of neighborhood boundaries demonstrate that individuals will demarcate their neighborhoods by using the physical and institutional characteristics of the neighborhood, its class, race, and ethnic composition, perceived criminal threats from within and outside the neighborhood, and symbolic neighborhood identities [2]. Guo and Bhat [17] offer a functional definition of neighborhoods as discreet,

non-overlapping communities generally defined by their physical geographies, natural advantages, and transportation systems. Urban neighborhoods have also been delineated by their patterns of land use such as types of residential or commercial buildings [9, 15].

Neighborhood delineation is a multidisciplinary task with two primary goals. The first is to define a neighborhood as a unit of observation. Neighborhoods are studied in housing market research [3], crime analysis [13, 29], and health research [25, 26, 31]. These studies are more concerned with understanding the effects a community and its built environment have on social processes rather than defining urban spaces with a common name. The second goal (and the primary aim of this paper) is more concerned with identifying the spatial extent of neighborhoods to populate gazetteers [21, 34, 35]. This has utility for queries in spatial databases. For example, a query to find all the restaurants within a particular neighborhood would rely on an accurate gazetteer containing the neighborhood boundaries.

The purpose of this research is to demonstrate a way to automate the delineation of urban neighborhood boundaries suitable for populating a gazetteer. A suitable method should draw on local data to delineate neighborhoods, allowing for an accurate representation of perceived boundaries. Additionally, the method should adhere to the logic that individuals use to demarcate their neighborhood boundaries. That is, the method should consider geographic context (e.g., major road boundaries or rivers) as a means to demarcate boundaries. Finally, the result of the method should account for spatial uncertainty in delineating the boundaries. Specifically, we apply a fuzzy-set model to address geographic vagueness as discussed in Varzi [32].

The following presents an expansion of the betashapes algorithm described in Erle [10]. While the original betashapes does not take into account context or uncertainty, it does provide a scalable solution to delineate neighborhood boundaries that adheres to the logic individuals use to demarcate their neighborhoods. We use the term μ-shapes to describe our algorithm because it incorporates fuzzy set theory to represent spatial uncertainty [36].

Fuzzy sets generalize traditional sets by generalizing the binary membership response, usually to a functional response with continuous range $[0, 1]$ [36]. Fuzzy sets and the accompanying fuzzy logic have been used in a broad range of residential and industrial (e.g., railroad systems) applications [33]. Fuzzy geographic applications were conceived early on [23], with the consequent fuzzy modeling of land-use classification [22, 8], and fuzzy representation of continuous spatial phenomena [14, 37].

This paper presents an expansion of the betashapes algorithm, incorporating fuzzy logic for the boundary demarcation process to quantify spatial uncertainty. In Section 2, we list the commonly used approaches to delineate urban neighborhoods. In Section 3, we present our method of delineating neighborhood boundaries. Section 4 presents an accuracy assessment for our method. Section 5 concludes with a discussion of future work.

2   Related work

A number of researchers [2, 6, 24, 30] have suggested using the aggregated cognitive maps of individuals to define neighborhood boundaries. In a study of neighborhoods in a city in Ohio, Coulton et al. [6] asked residents to draw their neighborhood boundaries onto a printed map. The researchers aggregated the core areas where 70% of respondents had drawn boundaries containing the same area [6]. Their research demonstrated that residents consistently drew their neighborhood boundaries along physical barriers or areas that were

considered to have different socio-demographic attributes. While this work demonstrates a way to map the perceived boundaries of urban neighborhoods, the time to recruit and train local experts is prohibitive in relation to setting provisions to collect volunteered geographic information.

Other solutions have been explored to derive the spatial footprints of vague geographic regions (e.g., neighborhoods) using aggregated volunteered geographic information [7, 16, 18]. These solutions rely on point-to-area methods to estimate the spatial footprint of imprecise geographic regions. One key advantage of point-to-area methods is that they are able to leverage volunteered geographic information (VGI) from popular social media platforms and thus provide a scalable means to populate gazetteers.

Kernel density estimates (KDE) are commonly used to delineate vague regions from point datasets [16, 20]. This method assigns a certainty to every point using a kernel function. This has the advantage of simplicity but lacks support for geographic context. Additionally, KDE allow for areas to be disjoint and may not create a realistic representation of neighborhoods which are compact and spatially contiguous.

Alani et al. [1] use Voronoi diagrams to estimate the areal extent of large vernacular regions of Scotland. This method requires the use of negative space. That is, it is necessary to have a set of points that are positively in the region as well as a set that are definitely outside of the region. A key limitation to this method is that the borders are unlikely to adhere to standard boundaries, since Voronoi cells are driven by the point distribution of the VGI.

Reinbacher et al. [27] used α-shapes to determine the spatial extent of the South of France from point coordinates. This approach cannot deal with context (e.g., physical boundaries) or areas with varying density. Density is especially problematic when using geographic information derived from social media posts. For example, popular tourist destinations tend to be over represented in Flickr datasets whereas residential areas are underrepresented in coverage [18]. This makes it difficult to determine the extent of primarily residential neighborhoods.

Wilske [34] uses geotagged social media documents (Flickr photos) annotated with place names to approximate the boundaries of neighborhoods in New York City. The author uses fuzzy sets to represent the spatial footprint of each neighborhood. The model represents a vague region through a pair of concentric regions with determinate boundaries [5]. This model assumes that a vague region has a core area that can be unambiguously assigned to the region, the lower approximation, surrounded by an area where membership within the region is uncertain, the upper approximation. Wilske [34] uses the spatial median derived from all geo-tagged documents to define the lower approximation of the model and defines the upper approximation as the convex hull (of the same set of geotagged documents).

The method described in Wilske [34] provides a scalable means to populate gazetteers using the spatial footprints of neighborhood boundaries. This method has two key limitations. First, the convex hull drawn around the individual VGI points is arbitrary and is unlikely to adhere to the perceived boundaries of individual residents. Second, the distance from the spatial median is not an informative measure for vague boundaries. The certainty that an area belongs to a neighborhood is not the same in all directions from the spatial median and is determined by other factors (e.g., physical barriers).

Erle [10] provides a possible point-to-area solution to generate neighborhood boundaries that considers the physical geography of a city. His method uses line datasets for

major physical barriers (e.g., roadways, rivers, railroads) to partition the area of a city into meaningful geographic primitives (blocks). Each block is then assigned to a neighborhood based on its proximity to a set of geographic coordinates tagged with a place name. A key limitation to this approach is that it does not account for spatial uncertainty when determining neighborhood boundaries. In the next section, we present a modification to the betashapes algorithm using an index to represent spatial uncertainty.

3 Methods

Given a spatial extent S and a set of associated features (e.g., roads, rivers, railways), we can partition S by the features' boundaries, into the set of blocks B. (While geographic feature data may not be sufficient to partition, e.g., a rural spatial extent or a spatial extent with few associated geographic layers, we assume the feature set of an urban environment, particularly one associated with VGI, is rich enough to partition S. Spatial extents with feature sets failing this assumption our beyond the scope of this work.) A μ-*shape* is a contiguous region $R \subseteq B$, such that, each $b \in B$ comprising R has one or more fuzzy metrics reflecting b's degree of membership to R. Intuitively, we think of each μ-shape as a fuzzy neighborhood. Given S and a set of polylines representing urban boundaries and natural barriers, point set P of VGI, and a user-defined number, k, of nearest neighbors (by Euclidean distance from the centroid of a block), the μ-shapes algorithm (Algorithms 1–3) generates a set of μ-shapes, each a contiguous collection of blocks, with boundaries that adhere to the physical structure of a city, and two fuzzy metrics. The first metric, $\mu_{\mathrm{VGI}}(b)$, is the fraction of VGI from block b's k nearest neighbors reporting membership to R. The second metric, $\mu_{\mathrm{ADJ}}(b)$, is the fraction of blocks adjacent to b, reporting the same parent neighborhood R as b.

Algorithm 1 μ-*shapes*

Input: A set of points P over a spatial extent S, each with a VGI neighborhood *name*; a set of polylines, representing urban boundaries and natural barriers; and a natural number k, where $1 \leq k \leq |P|$.

Output: A set of blocks B that partition S; for each block in B, the assigned parent neighborhood's *name* and μ_{VGI} and μ_{ADJ} values

1. Given S and a set of polylines, representing urban boundaries and natural barriers, generate the corresponding set of blocks B, where B is a partition of S.

2. For each block b in B:

    a. Determine the k nearest neighbors, the points closest to the centroid of b, P_b;

    b. *BlockNbhd*(b, P_b): compute b's parent neighborhood *name* and $\mu_{\mathrm{VGI}}(b)$;

    c. *BlockScore*(b, A_b): compute $\mu_{\mathrm{ADJ}}(b)$.

In this work, each block is assigned to a unique neighborhood, or in rare cases, unassigned. We could have defined the μ-shape to assign each block fuzzy metric scores for each neighborhood *name*, but this would have created a scaling challenge. Thus, we opted for low computational cost over complete fuzzy membership. Alternative definitions of

---

Algorithm 2 *BlockNbhd*(b, P_b)

Input: Block b and the k nearest points from the centroid of b, P_b
Output: The *name* of b's parent neighborhood and the VGI membership value of b, $\mu_{\mathrm{VGI}}(b)$

1. If *name* has a mode m among the points in P_b:
    a. Set block name to m, *b.name* $= m$;
    b. Compute $\mu_{\mathrm{VGI}}(b) = |m|/k$, the fraction of points in P_b with name *b.name*.
2. Else:
    a. If $|P_b| == 2$:
        i. Block b is not assigned to a neighborhood;
        ii. End.
    b. Else:
        i. Remove the furthest point from P_b, $P_b \leftarrow P_b -$ "furthest point";
        ii. *BlockNbhd*(b, P_b);
        iii. End.
    c. End.
3. End.

μ-shape could be formulated that enrich neighborhood membership with limited increase in computational cost.

μ_{VGI} and μ_{ADJ} address complementary aspects of neighborhood membership. The former metric is a measure of social consensus and therefore dependent on variations in the population that contributes VGI, as well as confounders, such as mobile device dead zones. The latter metric is a measure of physical adjacency and therefore subject to boundary effects and the geometric nature of the urban and natural boundaries. In the section that follows, we evaluate the μ-shapes algorithm and the performance of the two fuzzy metrics, via a case study.

Algorithm 3 *BlockScore*(b, A_b)

Input: Assigned block b and its adjacent blocks A_b
Output: The adjacency membership of b, $\mu_{\mathrm{ADJ}}(b)$

1. Initialize *Adj_Count* $= 0$.
2. For each adjacent block a in A_b:
    a. If *a.name* $==$ *b.name*:
        i. *Adj_Count*++;
    b. End.
3. $\mu_{\mathrm{ADJ}}(b) = Adj\_Count/|A_b|$, the fraction of blocks in A_b with the same name as b.
4. End.

---

4   Experiment and results

We present the performance of our algorithm against one technique for delineating neighborhood boundaries. Namely, we evaluate how well our algorithm compares to hand drawn boundaries by community experts. A full comparison of our algorithm to other methods is beyond the scope of this project and is left for future work.

The test dataset was derived from the Map Kibera Project (http://mapkibera.org), a resident led effort to map the neighborhoods and community amenities of the Kibera region of Nairobi, Kenya. The dataset consists of a shapefile of VGI, GPS points with an assigned toponym for one of fifteen neighborhoods. Additionally, the dataset contains a set of polygons for the neighborhood boundaries. The line dataset used to represent major physical barriers (and thus potential neighborhood boundaries) was derived from road, river, and railway data courtesy of OpenStreetMap, an open source repository of transportation data.

The Map Kibera Project is a community and NGO effort to collect significant geographic features of Kibera. The group splits into teams to collect points of interest (e.g., schools, clinics) throughout the community using GPS receivers. Each point has the neighborhood that the point resides in as part of its metadata. The neighborhood assignment to individual points was determined through consensus with the residents involved in the project. We use the point locations and the metadata as the input to test our algorithm. The neighborhood boundaries were hand-drawn over a satellite image of Kibera by a group of residents and volunteers. We use the resident drawn boundaries as the basis of comparison for our algorithm. The Kibera dataset was selected because the neighborhood boundaries were community derived rather than defined administratively or by a cartographic expert. Moreover, the VGI point data is well distributed and each labelled point falls into its respective neighborhood. Figure 1 shows the neighborhood boundaries overlaid with their respective neighborhood points.

Figure 2 shows the areal extent of Kibera overlaid with a dataset of aggregated line barriers (roadways, rivers, and railways). The blocks that are used for the evaluation were derived from the polyline data and the vector shapefile for the boundary of Kibera. The Kibera boundary data was split into pieces by the overlain polyline data. In the evaluation, the blocks are the spaces between the polylines. This was done in ArcGIS using the "Cut Polygons" tool [11]. The size and shape of the blocks are determined by, and sensitive to, the availability of, the polyline data.

Referring to the algorithm, each block is assigned to a neighborhood based on the mode nominal value of the nearest k VGI points. After each block is assigned to a neighborhood, the fuzzy metrics μ_{VGI} and μ_{ADJ} are computed, indicating the degree of membership of a block to the neighborhood. Figure 3 presents the neighborhood assignment. Each neighborhood derived from the μ-shapes algorithm is overlaid with its original neighborhood. Block misassignment can occur if either the original boundaries are not derived from a line barrier (e.g., a park) or if the point coverage for nearby primitives is sparse.

A key limitation of the original betashapes algorithm [10] is that it does not provide a measurement for spatial uncertainty. It is important to express and be able to visualize spatial uncertainty because of the inherent vagueness of neighborhood boundaries. Thus, the use of fuzzy sets in this paper. Figure 3 shows the neighborhoods derived from the μ-shapes algorithm overlaid with their respective blocks. The red boundaries illustrate the boundaries of the reference neighborhoods.
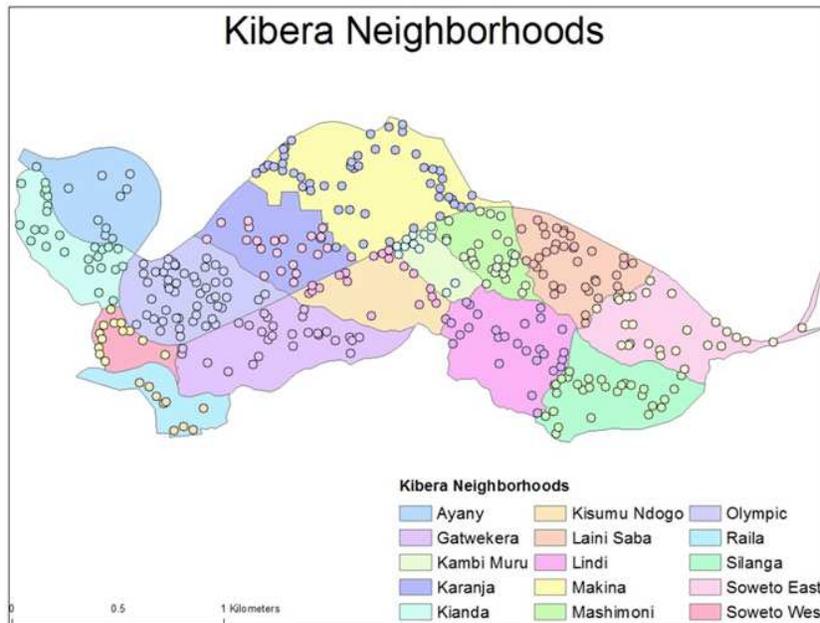
Figure 1: Kibera reference neighborhoods and the corresponding VGI point dataset, where each VGI point falls in its parent neighborhood.

Figure 4 presents the mapped μ_{ADJ} and μ_{VGI} results. The bottom panel presents neighborhood assignment using the adjacency fuzzy metric, μ_{ADJ}, while the top panel presents assignment when using the VGI fuzzy metric, μ_{VGI}. Here, we used Jenks natural breaks to stratify the symbology in each figure. Darker areas on the map indicate a higher membership value as determined by μ_{ADJ} or μ_{VGI}.

For comparison purposes with the reference dataset, which has a traditional set notion of membership, we merge the blocks by their assigned neighborhood, regardless of the fuzzy metric scores. In practice, a minimum fuzzy metric threshold may improve accuracy of neighborhood assignment. We opted to compare block assignment to neighborhoods against the reference dataset, using the minimum membership assignment.

To compare the algorithm's neighborhood assignment against the reference neighborhoods, a confusion matrix was calculated. The confusion matrix was derived by comparing the reference neighborhoods against the μ-shape neighborhoods (Figure 5). From the confusion matrix, a number of classification metrics can be derived. We computed the sensitivity, the accuracy, and the kappa coefficient for each neighborhood, and the overall accuracy and kappa coefficient for the algorithm's entire output. A kappa coefficient is a measure of inter-rater agreement which determines the extent to which the agreement of a classification is due to "true" agreement rather than mere "chance agreement" [4].

Tables 1 and 2 present the neighborhood and overall results, respectively. The overall accuracy of the neighborhood classification is approximately 90%, with a kappa coefficient of 91% better than random labelling. The average sensitivity is approximately 90% and all but four classes have a sensitivity greater than 85%. The lowest sensitivity (Kambi Muru)
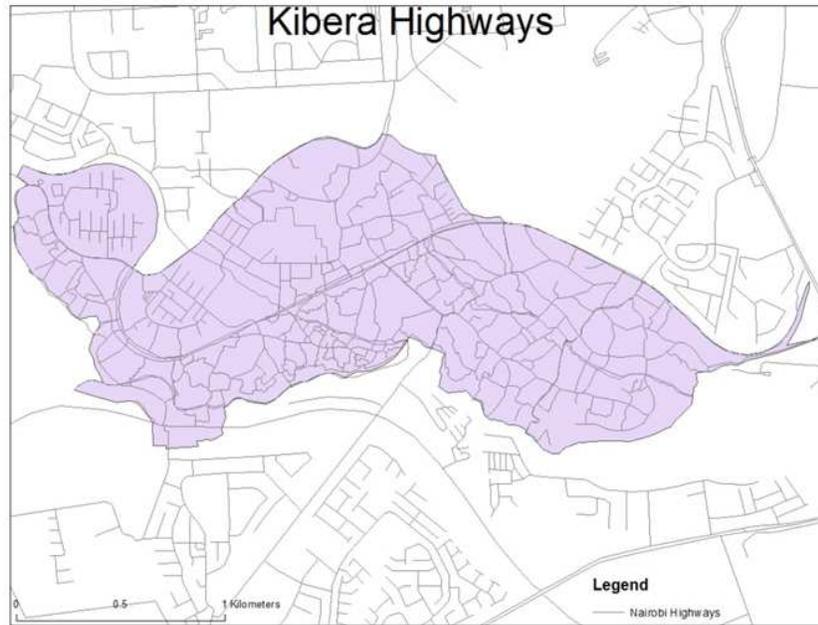
Figure 2: Kibera blocks ($n = 235$) generated by Kibera Highways line dataset from Open-StreetMap.

can be explained by the proximity measurement used to classify neighborhood blocks: while a VGI point labelled Kambi Muru overlaps a misclassified block, the mode for the nearest k points was attributed to the Kisumu Ndogo neighborhood. This demonstrates that the μ-shapes algorithm is sensitive to the selection of k and the VGI's distribution.

We evaluated μ_{ADJ} and μ_{VGI}'s ability to differentiate between a neighborhood's boundary blocks D (blocks that are adjacent to or overlap two or more neighborhoods), exterior blocks E (blocks that are not in D and on the boundary of the spatial extent S), and interior blocks I (blocks that are not in D or E), relative to the neighborhoods. To compute D, E, and I, we used a combination of spatial queries in ArcGIS [11], followed by visual inspection to finalize assignment to D, E, or I. This assignment was made only when there was agreement between the reference and computed neighborhoods and when a block could be assigned to a unique block category. From the 235 Kibera blocks, we found 99 boundary blocks D, 41 exterior blocks E, 71 interior blocks I, and 24 blocks for which assignment was not possible. We applied one-way analysis of variance test (ANOVA) to the values of μ_{ADJ} and μ_{VGI} by block group (D, E, and I), followed by multiple comparisons analysis (Table 3). For each fuzzy metric, for each block group pair, the table summarizes the difference of means, 95% confidence intervals of the lower and upper limits of the difference of means, and the p-value. If the difference of means interval does not contain 0 and $p \leq 0.05$, then the fuzzy metric has successfully differentiated between the block groups. We see that μ_{ADJ} and μ_{VGI} were able to distinguish between block groups D and E, and D and I, but not E and I. Since E and I can be distinguished by virtue of their relationship to

Assigned Blocks vs. Original Boundaries
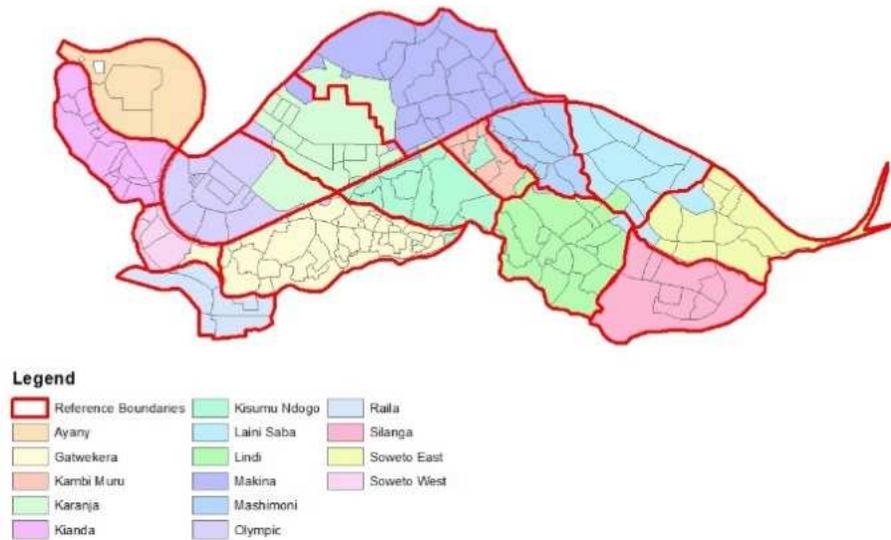
Figure 3: The reference Kibera neighborhoods (red polylines) and the algorithm's block assignment to neighborhoods (colored by neighborhood).

the boundary of the spatial extent S, the results imply that we can distinguish between any two block groups.

Since the ANOVA results are similar for μ_{ADJ} and μ_{VGI}, it is not clear that both metrics are needed to score block membership to a neighborhood. To assess the degree of independence of the two fuzzy metrics for this case study, we computed the linear correlation coefficient between μ_{ADJ} and μ_{VGI}, yielding $R^2 = 0.38$. The low correlation indicates a degree of independence between the metrics, suggesting that further investigation of μ_{ADJ} and μ_{VGI} is required.

5 Discussion and future work

The μ-shapes algorithm couples VGI data, specifically OpenStreetMap data and geotagged social media posts, with existing natural and infrastructural boundary data to address automated neighborhood delineation. Despite the data dependencies, in light of the recent surge in VGI, we expect the μ-shapes algorithm to be broadly applicable.

The confusion matrix comparison between the reference dataset and algorithm's output demonstrated high sensitivity and accuracy for this case study. ANOVA indicated that μ_{ADJ} and μ_{VGI} were able to distinguish between boundary (D) and interior (I) blocks, but neither metric could differentiate between exterior (E) and interior (I) blocks. However,
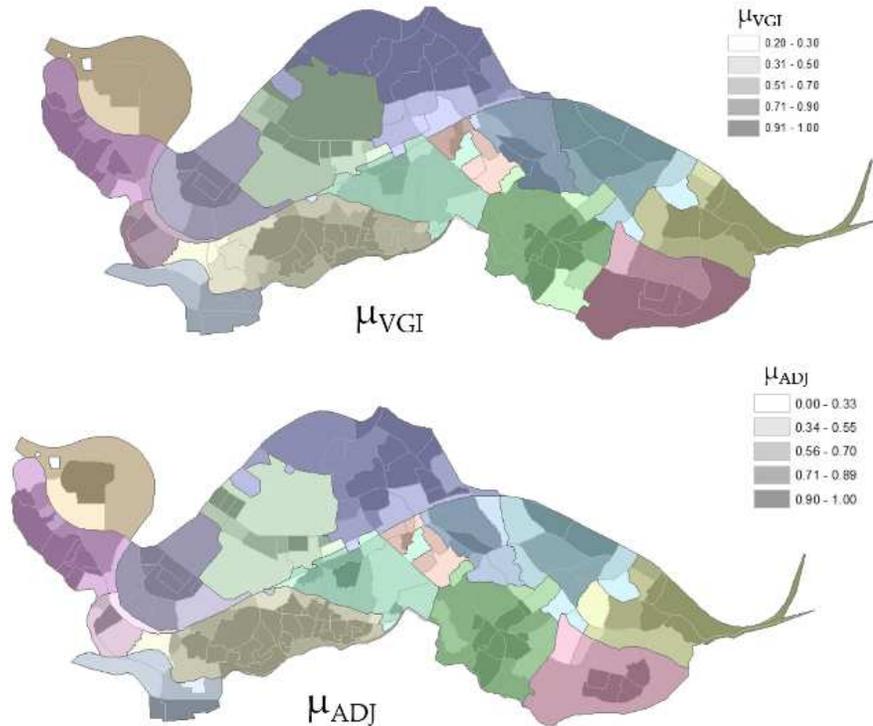
Figure 4: Kibera neighborhoods and the blocks comprising them, scored according to the fuzzy metrics μ_{ADJ} (bottom) and μ_{VGI} (top).

a simple spatial query can distinguish between blocks in E and I based on whether they lie on the boundary of the spatial extent S or not. Thus, given the inputs required by this research and a GIS to perform simple spatial queries, the μ-shapes algorithm proves effective in delineating neighborhoods, and in distinguishing between different parts of a neighborhood.

The μ-shape was defined for scalability, which is why each block is assigned to a unique neighborhood. For large geographic regions, it would be impractical to assess the degree of block membership to each neighborhood in the feature space. Having said that, alternative μ-shape definitions that score fuzzy block membership to multiple neighborhoods is conceivable, and may be desirable in certain instances.

There are a number of practical limitations to this algorithm. First, if a neighborhood lacks VGI representation, nearby neighborhoods will expand without limitation. Second, it requires supporting vector data to construct the blocks. Third, the algorithm performs poorly when there are too few points to initialize a mode or when the point distribution is biased. Finally, the output is sensitive to the selection of k (nearest neighbors).

It is worth noting that the Kibera dataset provides a near perfect dataset to test the μ-shapes algorithm. Other datasets, such as geotagged photos, will inevitably require some pre-processing step to filter outliers and mistagged points. The Kibera neighborhood boundaries were also mapped directly to hard physical barriers (i.e., roadways). The
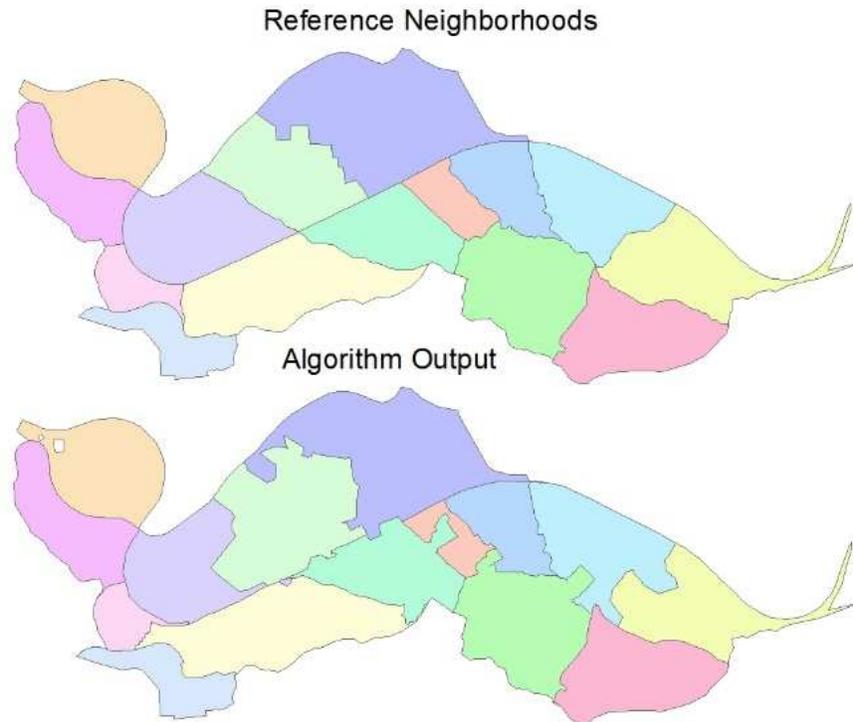
Figure 5: The reference Kibera neighborhoods (top) and the neighborhood assignments, as computed by the μ-shapes algorithm (bottom).

μ-shapes algorithm can accommodate other physical barriers, such as sharp changes in elevation or large parks, or more abstract boundaries, such as areas of different socio-economic status, only if these boundaries/barriers have been previously delineated.

Previous work has addressed topological classification of a region based on a point-set of data, and could be applied to handle misassignment arising from boundary, or other topological features [12, 19, 28]. This could support automated handling of each topological subcase rather than the present system that requires a human to form a judgement about the block's group membership, when the fuzzy metrics have low scores.

In future work, we will evaluate a broader scope of membership metrics, including the application of consensus over a set of metrics to arrive at a fuzzy determination of block membership to a neighborhood. The algorithm will also be generalized to accommodate sparse datasets, to weigh the importance of the k nearest neighbors according to their distance from the block, to let k vary according to the distribution of VGI, and to address regions where block computation is challenged by limited boundary information.

Acknowledgments

Reference class	Sensitivity	Accuracy	Class kappa
Kianda	1.00	0.97	1.00
Silanga	1.00	0.97	1.00
Ayany	0.98	1.00	0.98
Raila	0.98	0.98	0.98
Lindi	0.97	0.93	0.96
Kisumu Ndogo	0.96	0.86	0.96
Gatwekera	0.96	0.93	0.95
Laini Saba	0.95	0.87	0.94
Mashimoni	0.95	0.97	0.95
Soweto East	0.88	1.00	0.87
Karanja	0.87	0.62	0.86
Makina	0.85	0.95	0.83
Soweto West	0.78	0.92	0.77
Olympic	0.75	0.93	0.74
Kambi Muru	0.70	0.95	0.70

Table 1: Neighborhood-level assessment of the μ-shapes algorithm's classification.

	Overall Accuracy	Overall Kappa
[tbh!]	.90	.91

Table 2: Overall assessment of the μ-shapes algorithm's classification.

References

[1] ALANI, H., JONES, C. B., AND TUDHOPE, D. Voronoi-based region approximation for geographical information retrieval with gazetteers. *International Journal of Geographical Information Science 15*, 4 (2001), 287–306. doi:10.1080/13658810110038942.

[2] CAMPBELL, E., HENLY, J. R., ELLIOTT, D. S., AND IRWIN, K. Subjective constructions of neighborhood boundaries: lessons from a qualitative study of four neighborhoods. *Journal of Urban Affairs 31*, 4 (2009), 461–490. doi:10.1111/j.1467-9906.2009.00450.

[3] CLAPP, J. M., AND WANG, Y. Defining neighborhood boundaries: Are census tracts obsolete? *Journal of Urban Economics 59*, 2 (2006), 259–284. doi:10.1016/j.jue.2005.10.003.

[4] COHEN, J. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological bulletin 70*, 4 (1968), 213.

[5] COHN, A. G., AND GOTTS, N. M. The "egg-yolk" representation of regions with indeterminate boundaries. *Geographic objects with indeterminate boundaries 2* (1996), 171–187. 10.1.1.30.1000.

Metric μ_{ADJ}

Pair	LL 95% CI	Difference of Means	UL 95% CI	p
D, E	-0.45	-0.38	-0.30	9.56 E-10
D, I	-0.46	-0.40	-0.33	9.56 E-10
E, I	-0.10	-0.02	0.06	0.85

Metric μ_{VGI}

Pair	LL 95% CI	Difference of Means	UL 95% CI	p
D, E	-0.39	-0.31	-0.22	9.56 E-10
D, I	-0.39	-0.32	-0.25	9.56 E-10
E, I	-0.10	-0.01	0.07	0.95

Table 3: The results of balanced one-way ANOVA with post-hoc multiple comparisons analysis provides evidence that both μ_{ADJ} and μ_{VGI} differentiate between block groups D and E, and D and I, but not E and I.

[6] COULTON, C. J., KORBIN, J., CHAN, T., AND SU, M. Mapping residents' perceptions of neighborhood boundaries: A methodological note. *American Journal of Community Psychology 29*, 2 (2001), 371–383. doi:10.1023/A:1010303419034.

[7] DAVIES, C., HOLT, I., GREEN, J., HARDING, J., AND DIAMOND, L. User needs and implications for modelling vague named places. *Spatial Cognition & Computation 9*, 3 (2009), 174–194. doi:10.1080/13875860903121830.

[8] DRAGICEVIC, S. Coupling fuzzy sets theory and gis-based cellular automata for land-use change modeling. In *Proc. IEEE Annual Meeting of NAFIPS, Processing Fuzzy Information* (2004), vol. 1, IEEE, pp. 203–207. doi:10.1109/NAFIPS.2004.1336277.

[9] DUNCAN, M. J., WINKLER, E., SUGIYAMA, T., CERIN, E., LESLIE, E., AND OWEN, N. Relationships of land use mix with walking for transport: Do land uses and geographical scale matter? *Journal of urban health 87*, 5 (2010), 782–795. doi:10.1007/s11524-010-9488-7.

[10] ERLE, S. It's a beautiful day in the neighborhood! http://github.com/simplegeo/betashapes/blob/master/BLOG_POST.mkd, 8/5/102 2012.

[11] ESRI. Splitting a polygon by an overlapping feature. http://http://desktop.arcgis.com/en/arcmap/10.3/manage-data/editing-fundamentals/splitting-a-polygon-by-an-overlapping-line-feature.htm, 8/5/102 2012.

[12] FARAH, C., SCHWANER, F., ABEDI, A., AND WORBOYS, M. Distributed homology algorithm to detect topological events via wireless sensor networks. *Wireless Sensor Systems, IET 1*, 3 (2011), 151–160.

[13] FOSTER, K. A., AND HIPP, J. A. Defining neighborhood boundaries for social measurement: Advancing social work research. *Social work research 35*, 1 (2011), 25–35. doi:10.1093/swr/35.1.25.

[14] FOWLER, A., TEREDESAI, A. M., AND DE COCK, M. An evolved fuzzy logic system for fire size prediction. In *Proc. IEEE Annual Meeting of NAFIPS, Processing Fuzzy Information* (2009), IEEE, pp. 1–6. doi:10.1109/NAFIPS.2009.5156419.

[15] GREENBERG, S. W., AND ROHE, W. M. Neighborhood design and crime a test of two perspectives. *Journal of the American Planning Association 50*, 1 (1984), 48–61. doi:10.1080/01944368408976581.

[16] GROTHE, C., AND SCHAAB, J. Automated footprint generation from geotags with kernel density estimation and support vector machines. *Spatial Cognition & Computation 9*, 3 (2009), 195–211. doi:10.1080/13875860903118307.

[17] GUO, J. Y., AND BHAT, C. R. Modifiable areal units: Problem or perception in modeling of residential location choice? *Transportation Research Record: Journal of the Transportation Research Board 1898*, 1 (2004), 138–147. doi:10.3141/1898-17.

[18] HOLLENSTEIN, L., AND PURVES, R. Exploring place through user-generated content: Using flickr to describe city cores. *Journal of Spatial Information Science 1*, 1 (2010), 21–48. doi:10.5311/JOSIS.2010.1.3.

[19] JIANG, J., WORBOYS, M., AND NITTEL, S. Qualitative change detection using sensor networks based on connectivity information. *GeoInformatica 15*, 2 (2011), 305–328. doi:10.1007/s10707-009-0097-0.

[20] JONES, C. B., PURVES, R. S., CLOUGH, P. D., AND JOHO, H. Modelling vague places with knowledge from the web. *International Journal of Geographical Information Science 22*, 10 (2008), 1045–1065. doi:0.1080/13658810701850547.

[21] KESSLER, C., MAUÉ, P., HEUER, J. T., AND BARTOSCHEK, T. *Bottom-up gazetteers: Learning from the implicit semantics of geotags*. GeoSpatial semantics. Springer, 2009, pp. 83–102. 10.1007/978-3-642-10436-7_6.

[22] LEISS, I. A., SANDMEIER, S., ITTEN, K. I., AND KELLENBERGER, T. W. Use of expert knowledge and possibility theory in land use classification. In *Progress in Environmental Remote Sensing Research and Applicartions; Proceedings of the 15th EARSeL Symposium* (1995), pp. 133–137.

[23] LEUNG, Y. Fuzzy logic and knowledge-based gis. a prospectus. In *Proc. International Geoscience and Remote Sensing Symposium (IGARSS)* (1989), vol. 1, IEEE, pp. 47–50.

[24] LOHMANN, A., AND MCMURRAN, G. Resident-defined neighborhood mapping: Using GIS to analyze phenomenological neighborhoods. *Journal of Prevention & Intervention in the Community 37*, 1 (2009), 66–81. doi:10.1080/10852350802498714.

[25] MORENOFF, J. D. Neighborhood mechanisms and the spatial dynamics of birth weight. *American Journal of Sociology 108*, 5 (2003), 976–1017. doi:10.1086/374405.

[26] O'CAMPO, P. Invited commentary: Advancing theory and methods for multilevel models of residential neighborhoods and health. *American Journal of Epidemiology 157*, 1 (2003), 9–13. LR: 20111117; JID: 7910653; CON: Am J Epidemiol. 2003 Jan 1;157(1):1-8. PMID: 12505884; RF: 72; ppublish.

[27] REINBACHER, I., BENKERT, M., KREVELD, M. V., MITCHELL, J. S., AND WOLFF, A. *Delineating boundaries for imprecise regions*. Algorithms–ESA 2005. Springer, 2005, pp. 143–154. 10.1007/11561071_15.

[28] SADEQ, M. J., DUCKHAM, M., AND WORBOYS, M. F. Decentralized detection of topological events in evolving spatial regions. *The Computer Journal 56*, 12 (2013), 1417–1431. doi:10.1093/comjnl/bxs052.

[29] SAMPSON, R. J., AND RAUDENBUSH, S. W. Seeing disorder: Neighborhood stigma and the social construction of "broken windows". *Social psychology quarterly 67*, 4 (2004), 319–342. doi:10.1177/019027250406700401.

[30] TALEN, E., AND SHAH, S. Neighborhood evaluation using GIS—an exploratory study. *Environment and Behavior 39*, 5 (2007), 583–615. doi:10.1177/0013916506292332.

[31] TATLOW, J. R., CLAPP, J. D., AND HOHMAN, M. M. The relationship between the geographic density of alcohol outlets and alcohol-related hospital admissions in San Diego County. *Journal of Community Health 25*, 1 (2000), 79–88. doi:10.1023/A:1005145018975.

[32] VARZI, A. C. Vagueness in geography. *Philosophy & Geography 4*, 1 (2001), 49–65.

[33] WAKAMI, N., ARAKI, S., AND NOMURA, H. Recent applications of fuzzy logic to home appliances. In *Proc. International Conference on Industrial Electronics, Control, and Instrumentation (IECON)* (1993), IEEE, pp. 155–160. doi:10.1109/IECON.1993.339089.

[34] WILSKE, F. Approximation of neighborhood boundaries using collaborative tagging systems. *Proc. GI-Days* (2008), 179–187.

[35] WILSON, J. P., LAM, C. S., AND HOLMES-WONG, D. A. A new method for the specification of geographic footprints in digital gazetteers. *Cartography and Geographic Information Science 31*, 4 (2004), 195–207. doi:10.1559/1523040042742394.

[36] ZADEH, L. A. Fuzzy logic. *Computer*, 4 (1988), 83–93.

[37] ZHU, A.-X., HUDSON, B., BURT, J., LUBICH, K., AND SIMONSON, D. Soil mapping using GIS, expert knowledge, and fuzzy logic. *Soil Science Society of America Journal 65*, 5 (2001), 1463–1472. doi:10.2136/sssaj2001.6551463x.