Towards Fuzzy Geo-Set Visual Analysis

Romain Vuillemot*
LIRIS, École Centrale, Univ. Lyon

Liqun Liu†
LIRIS, École Centrale, Univ. Lyon

Théo Jaunet‡
LIRIS, Insa Lyon, Univ. Lyon

1 Context: Picking halfway locations in cities

Reasoning over spatial regions is frequent in everyday life. For instance, when moving to a new city, one may want to pick a locations based on its reachability by bus or at a walking distance to school and workplace. Geo-sets—spatial regions plus a category—are well suited to support picking such consensual locations: overlaps of two (or more) sets indicate places that match preferences (e. g., reachable by bus or close to schools). Figure 1 illustrates such sets intersections based on the transportation modality: flying like a bird, walking, or taking the bus. Those intersections may be complex especially as public transports distort time and space, resulting in blob-like shapes and fragmented areas. Also, as regions may be highly variable (e. g., bus timetables vary over the day, walking distance is approximate) their shape may expand or shrink over time (Figure 2). We refer to those changes using the umbrella term *variability* that may interchangeably capture time-varying parameters or uncertainty.

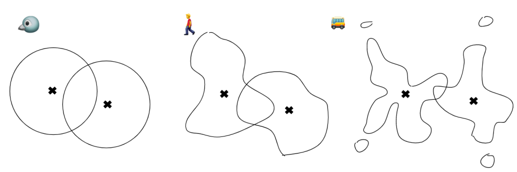

Figure 1: Examples of intersecting geo-sets.

This paper introduces a series of preliminary definitions and challenges related the visual analysis of such geo-sets intersections. Those have been raised from our experience working with geo-spatial analysis, in particular for the focus on analyzing *variability* of the sets (and consequently the variability of their intersections). The main task which we addressed in our work, and that requires better visual analysis tools, is to *"identify the location that minimizes travel time from two or three origins"* which is related to spotting the intersections with a specific set degree [3] (i. e. the one with maximum degree in our case).

2 Formal definitions

Let's consider P the space of all possible locations e, with $P = \mathbb{R}^2$. We define geo-sets S_i as sub-areas of this space $S_i \subset P$. Intersection between two sets $S_1 \cap S_2$ are elements belonging to both sets. Other intersections such as the $\emptyset$-intersection captures locations that do not fulfill any criteria, and the 1-intersection that only fulfill one. The generalization of this approach to n sets is S defined as follows $S = S_1, S_2, ..,$ with its intersection denoted $I(S)$ such as:

$$I(S) = \bigcap(S) = S_1 \cap S_2 \cap .. \cap S_j$$

We now introduce a *variability* factor to those geo-sets regions and intersections. In the traditional sets theory, elements either

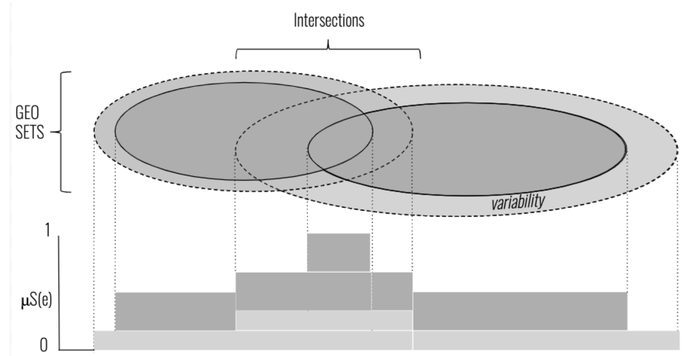

Figure 2: Intersection of geo-sets with variability (e. g., time) captured by the membership function $\mu_S(e)$.

belong or not to a set (using a crisp function such as $f_S(e) = 1$ if it does, 0 if it does not). We extend $f_S(e)$ into a membership function (inspired by Fuzzy Logic [9]) called $\mu_S(e)$ that assigns a confidence value to elements sets membership. This value is within $[0,1]$ and is defined as follows:

- 0 means that e belongs to no set at all (e. g., $\emptyset$-intersection)
- 1 means e belongs to the max-intersection (e. g., n-intersection)
- $0 < \mu_S(e) < 1$ means e belongs to at least one or multiple sets, but not all. Or at least not *fully* to all sets.

There are many ways to calculate $\mu_S(e)$, which usually is a domain- or data-specific method. Such calculation remains one of the main challenge we want to discuss during the workshop. Back to the geo-set related example of bus transit, $\mu_S(e)$ should capture the minimal and maximal coverage of the bus network over a given time period. For the walking distance geo-set, it should capture the boundaries of the minimal and maximal walking times. An approach to break-down this challenge of defining the global membership function, we define $w_S(e)$ as local weight function that captures local variability to each sets. Thus it allows us to formally defined $\mu_S(e)$ over the interval $[0,1]$ as the sum of all weighted memberships of element e:

$$\mu_S(e) = \frac{\sum_{S_i \in \{S_1,...,S_j\}} w_{S_i}(e)}{|S|}$$

The intuition behind this calculation (illustrated Figure 2) is that it assigns smaller value to locations e belonging to the periphery of geo-sets with high variability.

3 Challenges

Challenge 1: Complexity of sets intersection

As a direct consequence of adding variability, new geo-sets are added (the partial ones with $w_S(e)$ between 0 and 1) thus the number of intersections explodes. Using scalable methods such as [2, 8] would still face limits as the number of sets grows polynomially, and also they would loose geo-context as they do not represent sets on a 2D space. Area-based sets drawing methods [1, 5] would also gain in visual complexity if variability is encoded (e. g., using opacity or additional color scales).

*e-mail: romain.vuillemot@ec-lyon.fr
†e-mail: liqun.liu@ec-lyon.fr
‡e-mail: theo.jaunet@insa-lyon.fr

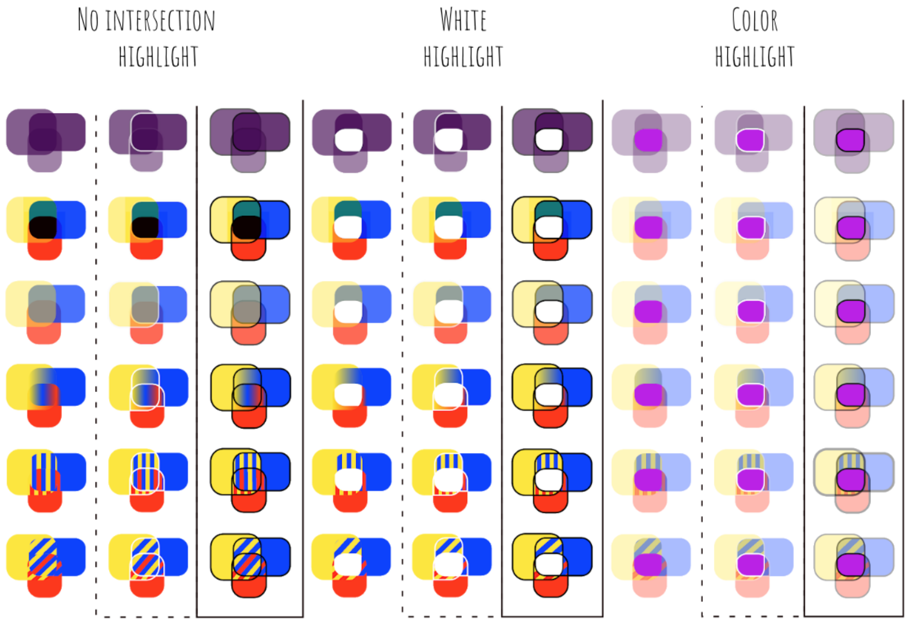

Figure 3: Examples of geo-sets intersections design variations. Rows illustrate design variations by: opacity, primary colors, primary colors and opacity, shading, vertical and horizontal hatching. Columns illustrate intersections design variations.

Challenge 2: Revisiting set degrees and intersections tasks

For crisp sets, elements degree is the number of sets it belongs to. Fuzzy geo-sets sets intersections redefine this concept, has they have a degree plus a membership value. Thus additional tasks are introduced, such as seeking for sets with a specific membership $\mu_S(e)$ value (e. g., 50%). Also, cardinality (number of elements in a set) is impacted as sets membership is continuous, thus seeking for specific elements also comes with matching elements *and* matching with a membership value.

Challenge 3: Time-varying data

Real-life geo-related dataset constantly change over time. As introduced earlier, the variability aims at capturing geo-sets size changes. But in the meantime, elements may vary in location over time. Previously introduced measures of intersections and their elements may be tied to a specific time period, and may vary over time. We acknowledge this challenge is not specific to fuzzy geo-sets, but we found it important to support it with real-life geo-located datasets.

4 Ongoing work and perspectives

As far as we know those challenges are still open for the sets visual analytics community. We conducted preliminary work regarding the visual encoding of $\mu_S(e)$ and the geo-sets intersections. Figure 3 illustrates preliminary design investigations related to color blending as an approach to multi-class representations [7], but that emphasize intersections. The rationale behind this work is to focus on a generative approach to explore design variations, and let users pick relevant combinations.

More efforts regarding fuzzy geo-sets encoding should be conducted. The Rose Diagrams [4] provides a first attempt to convey membership values, but lacks of geo-references like the ones that provide Line Set [1], Kelp Diagrams [5], which in return to not convey membership values. As a final note, while we introduced the definitions and challenges for mobility analysis in city, similar challenges exist for other domains such as images analysis [6] that may require to support fuzzy regions comparison.

5 Acknowledgment

We would like to thank the organizers of the workshop and the reviewers for their very thoughtful comments.

References

[1] B. Alper, N. Riche, G. Ramos, and M. Czerwinski. Design Study of LineSets, a Novel Set Visualization Technique. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2259–2267, Dec. 2011. doi: 10.1109/TVCG.2011.186

[2] B. Alsallakh, W. Aigner, S. Miksch, and H. Hauser. Radial Sets: Interactive Visual Analysis of Large Overlapping Sets. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2496–2505, Dec. 2013. doi: 10.1109/TVCG.2013.184

[3] B. Alsallakh, L. Micallef, W. Aigner, H. Hauser, S. Miksch, and P. Rodgers. *Visualizing Sets and Set-typed Data: State-of-the-Art and Future Challenges*. The Eurographics Association, 2014. doi: 10.2312/eurovisstar.20141170

[4] A. R. Buck and J. M. Keller. Visualizing uncertainty with fuzzy rose diagrams. In *2014 IEEE Symposium on Computational Intelligence for Engineering Solutions (CIES)*, pp. 30–36, Dec. 2014. doi: 10.1109/CIES.2014.7011827

[5] K. Dinkla, M. J. v. Kreveld, B. Speckmann, and M. A. Westenberg. Kelp Diagrams: Point Set Membership Visualization. *Computer Graphics Forum*, 31(3pt1):875–884, 2012. doi: 10.1111/j.1467-8659.2012.03080.x

[6] H. Haußecker and H. R. Tizhoosh. Fuzzy Image Processing. In *Computer Vision and Applications*, pp. 541–576. Elsevier, 2000. doi: 10.1016/B978-012379777-3/50017-0

[7] J. Jo, F. Vernier, P. Dragicevic, and J.-D. Fekete. A Declarative Rendering Model for Multiclass Density Maps. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):470–480, Jan. 2019. doi: 10.1109/TVCG.2018.2865141

[8] A. Lex, N. Gehlenborg, H. Strobelt, R. Vuillemot, and H. Pfister. UpSet: Visualization of Intersecting Sets. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1983–1992, Dec. 2014. doi: 10.1109/TVCG.2014.2346248

[9] L. A. Zadeh. Fuzzy sets. *Information and Control*, 8(3):338–353, June 1965. doi: 10.1016/S0019-9958(65)90241-X