Quantitative Platial Analysis: Methods for Handling and Representing Platial Heterogeneity and Linking Varying Concepts of Place

– Invited Keynote Paper –

Alexis Comber, Alice Butler, Nick Malleson, and Alex Schafran

School of Geography, University of Leeds, UK

This paper explores potential approaches for quantitative platial analysis. It revisits some of the early work examining place in social media data in light of recent proposals for a platial GIS. Focussing on Massey's concept of space that incorporates a sense of belonging and kinship, where space becomes place through social relations, it uses coded Twitter data containing the term "shithole" to generate a predictive models of different types of platial denigration. These are used to infer the spatial distribution of different types of platial denigration. The results show that there is little spatial pattern to denigration of different places and sports facilities, but that denigration of one's own local area and of one's own personal space have highly localized distributions. The discussion indicates a number of areas for further research with a particular warning against developing platial GISs as has been suggested by many authors. Other explicitly GIScience avenues may be more productive and insightful.

Keywords: spatial analysis; platial analysis; Twitter; shithole

1 Introduction

This introduction briefly covers the concept of place in geography and then the inherent social construction of spatial data in more detail. These lay the ground for a critique of how the GIScience/spatial analysis/geocomputation community have hitherto sought to take a platial turn, and set up the platial analysis later in the paper.

The concept of place is a core consideration in critical geography. It has a number of characteristics that GIScientists have struggled to robustly accommodate within a *platial* information system: place refers to multiple spatial concepts; places are spaces where the notion of distance is irrelevant; and place defines the socio-cultural context in which everyday lives are lived. Doreen Massey, offers a useful conceptualization of the idea of "place", that emphasizes the changing nature of place and place-making (Massey, 2000). She clearly and concisely defines what constitutes a place: "places are spaces of social relations" (Massey, 2000, p. 459). This definition highlights the key difference between space and place and highlights the centrality of social relations for place-making. In this the concept of "place" incorporates a sense of belonging and kinship where space is anonymous but has the potential – with the introduction of social relations – to become place. Places also evolve and are dynamic: "the place goes on being made" (Massey, 2000, p. 464) emphasizing the force of time and the necessity of understanding the entirety of a place rather than at a "snapshot" moment in time. Place is also relative to the multiple and differentiated and public: "'one place' can be known in numerous ways" (Massey, 2000, p. 464), suggesting that, in the same way that time changes a place so, too, does one's social

A Comber, A Butler, N Malleson, and A Schafran (2018): *Quantitative Platial Analysis: Methods for Handling and Representing Platial Heterogeneity and Linking Varying Concepts of Place*. In: R Westerholt, F-B Mocnik, and A Zipf (eds.), Proceedings of the 1st Workshop on Platial Analysis (PLATIAL'18), pp. 7–14

https://doi.org/10.5281/zenodo.1472735

PLATIAL'18 First Workshop on Platial Analysis (PLATIAL'18)
Heidelberg, Germany; 20–21 September 2018

relations with that place. There is a strong nostalgic element to this concept of place, and for Massey impersonal, symbolically distant "space" becomes "place" when social relations and memories are interwoven with physical space. Thus place-making is ongoing and relates to belonging. A place is known and intimate and depends on the imposition of bodies and relations between those bodies in order to exist as anything other than space. Massey's formulation of space and place offer to this study an understanding that foregrounds the process of place: place is ongoing and continually shaped.

How we represent the real world in our spatial databases is a key concept in geography as it determines the nature of the questions we are able to answer in our geographical and spatial analyses. Consequently, notions of place have also been extensively considered within the broad domain of GIScience and a long-standing corpus of research exists. As GIS started to emerge into the mainstream in the early 1990s, many researchers started to critically examined how were using the technology and digital data, particularly how real world features were delineated and encoded in spatial databases. The work of Barry Smith, David Mark, and Andrew U. Frank are exemplars in this area. They were concerned with the concepts and meaning that are implicitly embedded in data, how features were delineated, their labels ("lake" vs. "lac"), the semantics, culture and philosophies they represented, and how to appropriately encode them in our databases. There are three key and interlinked but forgotten research ideas from this time that worth revisiting:

1. The contested nature of features through the notion of fiat and bone fide objects and boundaries (Burrough and Frank, 1996; Smith, 1995, 2001; Smith and Varzi, 2000). In brief, *fiat* objects exist only because of some kind of cognitive demarcation and owe their existence to acts of human decision. *Bona fide* objects do not and are independent of human conceptions. An example *fiat* might be representational choices over where to place the forest boundary as trees intergrade with shrub land cover in successional vegetation. Such choices are routinely made in the construction of all spatial data and inevitably have the potential to result in analytical variation and therefore uncertainty;

2. Acknowledgement that different groups of people conceptualize the world in different ways. The names and labels we give to things, places and geographic phenomena reflect group perceptions of characteristics (Mark and Turk, 2003; Smith and Mark, 2003) due to linguistic and cultural factors (Smith and Mark, 1998). This recently been observed in crowdsourced data (Comber et al., 2016) and data from different groups has been shown to result in significantly different results when used in analysis; and

3. Geographic objects or processes and their group meanings also vary fundamentally with scale (Fisher et al., 2004).

In the early days of GIS/GIScience researchers were fundamentally concerned with these core representational issues and the uncertainties that might occur when, e. g., data and user perceptions of an object differ. These considerations persist and imply that spatial data will always be socially constructed (Harvey, 2000). They also lead to a health warning that has largely been ignored as the geo-digital revolution: geographic entities are inherently and intimately connected to the space that they occupy and to the manner of their (human) conceptualization (Varzi, 2001).

These issues, fundamental to all spatial data and for spatial data analysis, have largely been overlooked by the GIScience community (COSIT excepted) in recent years. This is mostly due to the nature of digital information systems and the ease now with which we are able to collect, process, and analyse spatial data of all kinds. Our situation is analogous to the old joke "What is a lecture?"[1].

It is also reflected in recent forays (like this one) considering how GIScience and other information technologies could take a platial turn. These have been driven by the opportunities for digital place-based research afforded by technological developments. A number of papers have generally made the following points (extracted variously from text) (Gao et al., 2013; Goodchild, 2011, 2015; Roche, 2016):

1. Places are messy and difficult to define and pin down. Places are poorly defined in "space", frequently with indeterminate boundaries, and the individual perceptions of those places and their properties vary. However, for GIS they need to be "identifiable" to exist and consequently named places, despite frequently being vaguely defined and context dependent, are used to provide the link between (Euclidian, mapped) space and behavioural place, because place-names can be converted to coordinates.

2. GIS is not very good at representing place. Although GI technologies are inherently spatial, they can be used as a mediating object in planning and to support citizen engagement. But they are not very good at handling alternative (i. e., non-Euclidean) or ambiguous representations of places. This is because places within any human discourse may be vaguely defined and context dependent in contrast to the precise and objective coordinates of space.

3. Actually, beyond understanding behaviours through place names we do not know how to do this. New personal digital GPS-enabled technologies provides opportunities for a new relationship between GIScience and place through social media, VGI, geoweb, etc. This would support spatial enablement, literacy, and empowerment but new theories are needed for such geoplatial methods, technics, and tools.

Integrating platial and spatial in this manner may be mis-guided. They can certainly be linked (see the analysis below) but each has their unique strictures and rubrics. Robust theories take time to evolve, the technology is moving faster than the thinking, and we still have basic things that need to be addressed in GIScience (e. g., we can't even deal with time very well). Spatial detective work is at the core of the role spatial analysis/GIScience: our job is to help domain experts understand what is going in their place (and our space) by developing methods for (s)p(l)atial explorations of data and processes.

There alternatives to the "we need a new philosophy-geoplatial turn" route: returning to previous research we can see that many platial-spatial paths that we have been discarded without being exhausted – mainly because of the technological and data neophilia that has pervaded information sciences in recent years. One such direction is the work by GIScientists in the early 2000s that explicitly sought to link to descriptions and mappings of place to construct alternative gazetteers, vernacular geographies, to create maps of "places", etc. using text and sentiment mining of place names in Flickr tags, alternative POIs, and citizen participatory mapping. New forms of data were becoming available and provided opportunities to understand how people were experiencing their environment. This activity acknowledged the fact that notions of place *are* grounded semantics, meaning, (spatial) cognition, perception, and linguistics, which researchers sought to capture. It also acknowledged the inherent relativity of place and sought to construct multiple geographies. It recognized that any given palatial concept may be understood differently by different people, groups, and cultures. Such approaches provide a framework (rather than a philosophical and ontological tautology) for GIScientists to work with platial data. It is illustrated in the next section through a spatial analysis of where people use the term "shithole" in relation to different "places".

2 Data and Methods

Butler et al. (2018) examined the use and intended meaning of the term "shithole" in tweets to understand place based stigma and how discourses of denigration are shaped by the availability and uptake of social media platforms. They found that in most cases, "shithole" or "#shithole" was used to refer to places that the tweeter was not from and that some uses reflected a desire to leave (predominantly contributed by female users). The geography of these tweets had no specific pattern, leading the authors to note that "individuals cry for help and want to leave virtually everywhere and every type of place. They want to leave dirty, ill-equipped homes, villages that are boring, towns that lack amenities, and cities that are dirty and full of 'others'" (Butler et al., 2018, p. 11).

This paper uses Butler's data to create predictive spatial models. The data included the following relational geography codes to indicate the scale, scope, and target of each tweet:

- **Other** when the tweet referred to a place that was not the home of the tweeter;
- **Own** where the tweet referred to the tweeter's own area or region;
- **Facilities** such as a sports stadium;
- **Personal** which usually was used to refer to a room, home, or place of work.

The 1989 tweets containing "shithole" were stemmed, creating a corpus of 2653 stemmed terms and used to train an elastic net/lasso (ELN) model. The model was then applied to a larger, uncoded,

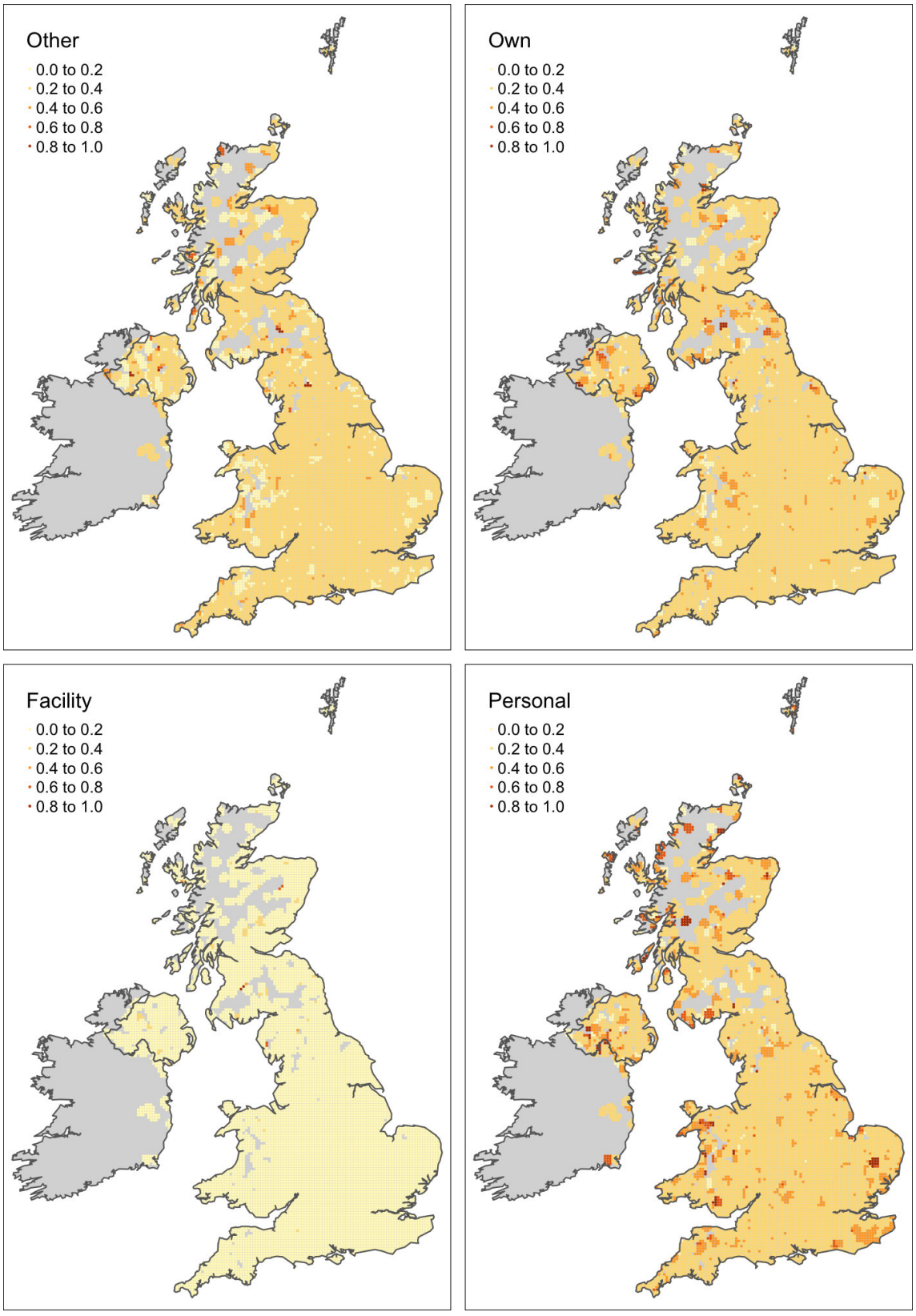

Figure 1: The spatial distribution of tweets classified into the different relation geography classes of "shithole": the mean posterior probability of each class.

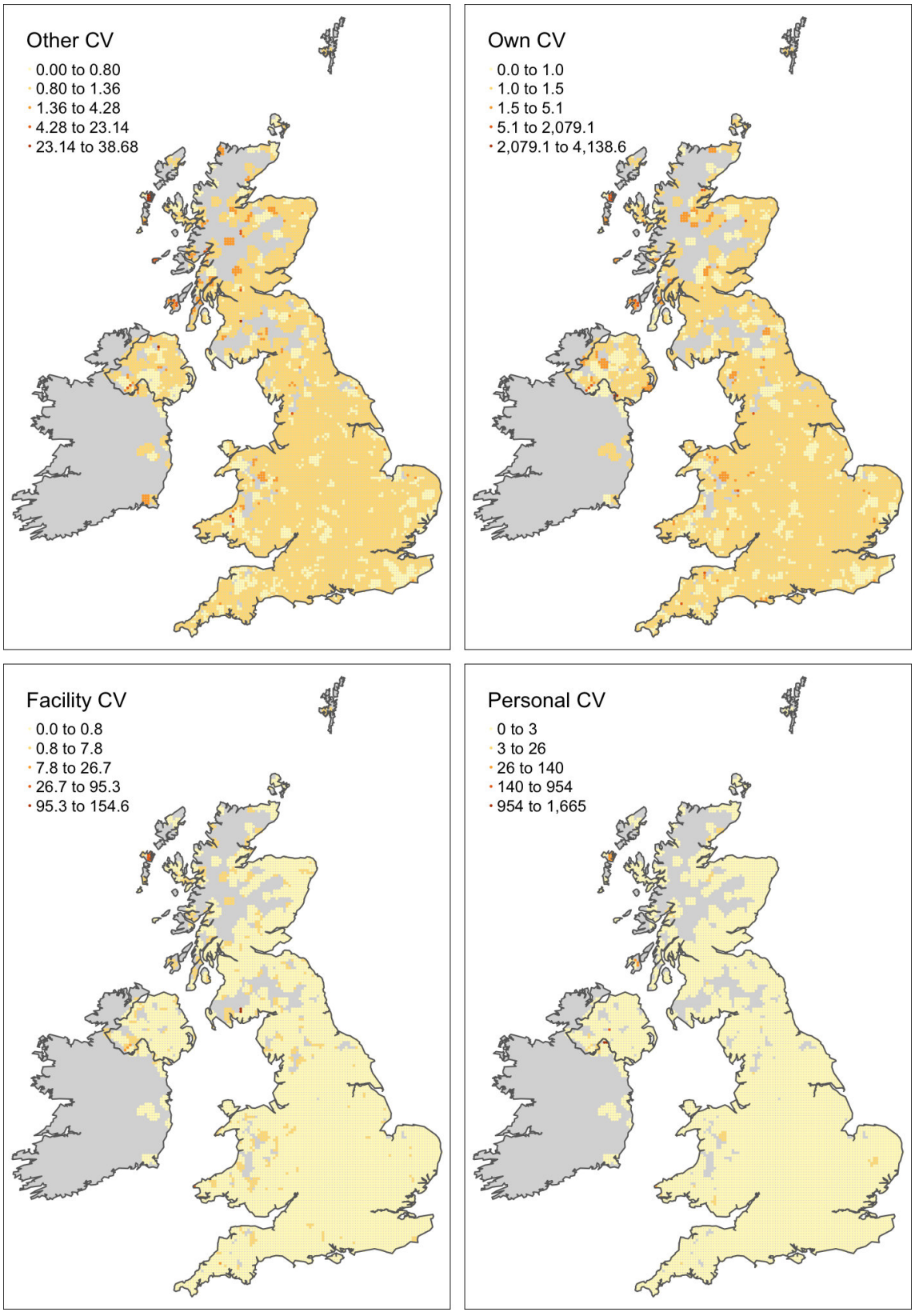

Figure 2: The spatial variation of the mean posterior probability of each class.

dataset of approximately 1 million geo-located (not geo-tagged) tweets and used to predict the relational geography class. The spatial distribution of each class forms the basis of the analysis reported in this paper. The aim, however, was not just to generate a hard, crisp classification of tweets, but to explore the underlying geographies of variations in the strength of classification through the model posterior probabilities and their variances. These soft classification measures provide a route to understand spatial variation in the way that the term "shithole" is used in different parts of the country.

3 Results

The ELN models was constructed from a binary sparse matrix of stemmed terms, indicating the presence of a given term. In-sample model fits showed the model to correctly predict the class of "shithole" 94.3% of the time. The model was then used to predict the 4 classes for each of the 961,597 tweets. These were summarized and smoothed over a 5 km grid using a 10 km radius circular window. For each location, the mean posterior probabilities for each class from the ELN model were calculated along with the coefficient of variation. The first gives an indication of the general trend and the second indicates the heterogeneity of that value. The maps of these are shown in Figures 1 and 2 for each class.

The maps in Figure 1 show a number of things, some of which were identified by text (Butler et al., 2018):

- **Other** has no discernible geographic pattern indicating that tweeters everywhere are equally likely to denigrate a different place.
- **Own** has strong local concentrations in the Scottish borders, Northern Ireland and Mid-Wales indicating higher levels of expressions of denigration by residents in these areas of their own locality than others.
- **Facility** has an even distribution as Butler stated – simply, there is no geography.
- **Personal** is heavily found at the geographic fringes in rural remote places. This suggests that in these places, tweets tend to be closer to denigration of personal space (room, home, or place of work) than any other of the classes. People do not like their remote lives and social media gives them an opportunity to express that.

The maps in Figure 2 indicate that the variation in pooled posterior probability is relatively even for each of the four classes, with a few notable exceptions:

- **Other** has some notable pockets of high variation in the Hebrides and Grampians in Scotland, Cork in Ireland, and Mid Wales.
- **Own** has similar localized trends to **Other**.
- There are no obvious trends in the variation of **Faculty** or **Personal**.

Places with low variation in Figure 2 indicate that the mean value in Figure 1 is highly representative of the tweets in that area.

4 Discussion

This research used a standard technique to model spatial variation in perceptions of place, and by developing a predictive model from data that had been manually coded for platial characteristics. This was applied to other Twitter data, for which the geography was known. The geography was derived from the Twitter users' home location and not the target of their tweets. So, e. g., areas with high posterior probabilities for "Other" are not locations that are denigrated by people who do not live there. Rather it indicates a greater probability for people in that region being more likely to denigrate other places in their tweets.

The analysis used a 5 km grid to summarize the typical values of the classified tweets and their variations. Summarizing and visualizing data in this way provides a useful starting point for discussions

with domain experts – there may be well known social gradients that are described by these mapped distributions. The mean posterior probabilities provide an indication of the multiple potential places that may be present at any location. Soft representations such as these (and, e. g., fuzzy sets) allow alternative and multiple representations of the same geographic phenomena, ones that allow different perceptions of space and thereby place to be accommodated (e. g., Comber and Kuhn 2018).

Finally, there are lots of areas for further work that will be expanded in the full journal paper arising from the publication: alternative classification models (initial work showed quite different results for "Personal" using linear discriminant analysis rather ELN), the use of medians and inter-quartile ranges rather than means and coefficients of variation to quantify central tendencies in a way that is resistant to numerical outliers, deeper investigation of specific locales to try to unpick and understand the local trends that have been observed here, and testing the sensitivity of the processes captured in the predictive models to scale and how the patterns and trends observed vary over different units of analysis. Finally it would be useful to evaluate the representativeness of the platial phenomena captured by the classified twitter data from some alternative data source.

This kind of research avenue is likely to be more productive and will better support platial analysis than ones that seek to steer GIScience towards a platial turn: we should stick to what we are good at – representation, scale, and uncertainty – areas that no other disciplines can do as well as us.

Notes

1. Answer: It is the process by which the lecturer's words are transferred to the student's notes without going through the brain of either.

ORCID

Alexis Comber https://orcid.org/0000-0002-3652-7846
Alice Butler https://orcid.org/0000-0002-7205-9832
Nick Malleson https://orcid.org/0000-0002-6977-0615
Alex Schafran https://orcid.org/0000-0003-1990-925X

References

Burrough, Peter A and Frank, Andrew U (eds.): *Geographic objects with indeterminate boundaries*. London, UK: CRC, 1996

Butler, Alice; Schafran, Alex; and Carpenter, Georgina: *What does it mean when people call a place a shithole? Understanding a discourse of denigration in the United Kingdom and the Republic of Ireland*. Transactions of the Institute of British Geographers, 43(3), 2018, 496–510. doi: 10.1111/tran.12247

Comber, Alexis and Kuhn, Werner: *Fuzzy difference and data primitives: a transparent approach for supporting different definitions of forest in the context of REDD+*. Geographica Helvetica, 73(2), 2018, 151–163. doi: 10.5194/gh-73-151-2018

Comber, Alexis; Mooney, Peter; Purves, Ross S; Rocchini, Duccio; and Walz, Ariane: *Crowdsourcing: it matters who the crowd are. the impacts of between group variations in recording land cover*. PloS ONE, 11(7), 2016, e0158329. doi: 10.1371/journal.pone.0158329

Fisher, Peter; Wood, Jo; and Cheng, Tao: *Where is Helvellyn? fuzziness of multi-scale landscape morphometry*. Transactions of the Institute of British Geographers, 29(1), 2004, 106–128. doi: 10.1111/j.0020-2754.2004.00117.x

Gao, Song; Janowicz, Krzysztof; McKenzie, Grant; and Li, Linna: *Towards platial joins and buffers in place-based GIS*. Proceedings of the 1st ACM SIGSPATIAL International Workshop on Computational Models of Place (COMP), 2013, 42–49. doi: 10.1145/2534848.2534856

Goodchild, Michael F: *Formalizing place in geographic information systems*. In: Burton, Linda M; Matthews, Stephen A; Leung, ManChui; Kemp, Susan P; and Takeuchi, David T (eds.), *Communities, neighborhoods, and health: expanding the boundaries of place*, New York, NY: Springer, 2011. 21–34. doi: 10.1007/978-1-4419-7482-2_2

—— *Space, place and health*. Annals of GIS, 21(2), 2015, 97–100. doi: 10.1080/19475683.2015.1007895

Harvey, Francis: *The social construction of geographical information systems*. International Journal of Geographical Information Science, 14(8), 2000, 711–713. doi: 10.1080/136588100750022741

Mark, David M and Turk, Andrew G: *Landscape categories in Yindjibarndi: ontology, environment, and language*. Proceedings of the 6th International Conference on Spatial Information Theory (COSIT), 2003, 28–45. doi: 10.1007/978-3-540-39923-0_3

Massey, Doreen: *Living in Wythenshawe*. In: Borden, Iain; Kerr, Joe; Rendell, Jane; and Pivaro, Alicia (eds.), *The unknown city. Contesting architecture and social space*, Cambridge, MA: MIT Press, 2000. 458–474

Roche, Stéphane: *Geographic information science II: less space, more places in smart cities*. Progress in Human Geography, 40(4), 2016, 565–573. doi: 10.1177/0309132515586296

Smith, Barry: *On drawing lines on a map*. Proceedings of the 2nd International Conference on Spatial Information Theory (COSIT), 1995, 475–484. doi: 10.1007/3-540-60392-1_31

—— *Fiat objects*. Topoi, 20(2), 2001, 131–148. doi: 10.1023/A:1017948522031

Smith, Barry and Mark, David M: *Ontology and geographic kinds*. Proceedings of the 8th International Symposium on Spatial Data Handling (SDH), 1998, 308–320

—— *Do mountains exist? towards an ontology of landforms*. Environment and Planning B, 30(3), 2003, 411–427. doi: 10.1068/b12821

Smith, Barry and Varzi, Achille C: *Fiat and bona fide boundaries*. Philosophical and Phenomenological Research, 60(2), 2000, 401–420. doi: 10.2307/2653492

Varzi, Achille C: *Introduction*. Topoi, 20(2), 2001, 119–130. doi: 10.1023/A:1017944405193