Interdisciplinary Perspectives on Place

Proceedings of the
Second International Symposium
on Platial Information Science

PLATIAL'19

5–6 September 2019
Coventry, UK

Franz-Benjamin Mocnik and René Westerholt
(editors)

Please Cite This Volume As

FB Mocnik and R Westerholt (eds., 2020): *Interdisciplinary Perspectives on Place. Proceedings of the 2nd International Symposium on Platial Information Science (PLATIAL'19)*

https://doi.org/10.5281/zenodo.3628833

Location and Dates
Coventry, United Kingdom; 5–6 September 2019

Convenors
René Westerholt (University of Warwick, United Kingdom)
Franz-Benjamin Mocnik (University of Twente, the Netherlands)

Keynote Speakers
Nigel Thrift (University of Oxford, United Kingdom)
Thora Tenbrink (Bangor University, United Kingdom)

Programme Committee
Thomas Blaschke (University of Salzburg, Austria)
Dirk Burghardt (Technical University of Dresden, Germany)
Alexis Comber (University of Leeds, United Kingdom)
Clare Davies (University of Winchester, United Kingdom)
Ekaterina Egorova (Massey University, New Zealand)
Sara I Fabrikant (University of Zurich, Switzerland)
Michael F Goodchild (University of California, Santa Barbara, United States)
Krzysztof Janowicz (University of California, Santa Barbara, United States)
Karen Kemp (University of Southern California, Dornsife, United States)
Grant McKenzie (McGill University, Canada)
Franz-Benjamin Mocnik (Heidelberg University, Germany)
Alenka Poplin (Iowa State University, United States)
Simon Scheider (Utrecht University, the Netherlands)
Kathleen Stewart (University of Maryland, United States)
Thora Tenbrink (Bangor University, United Kingdom)
Nigel Thrift (University of Oxford, United Kingdom)
Maria Vasardani (RMIT University, Australia)
René Westerholt (University of Warwick, United Kingdom)
Stephan Winter (University of Melbourne, Australia)

EDITORIAL

INVITED PAPER

PLACE AND HUMAN GEOGRAPHY

PLACE AND PSYCHOLOGY

PLACE AND SOCIAL MEDIA

Introduction to the Second International Symposium on Platial Information Science

– Editorial –

Rene Westerholt[1,2] and Franz-Benjamin Mocnik[3]

[1]*School of Spatial Planning, TU Dortmund University, Germany*
[2]*Centre for Interdisciplinary Methodologies, University of Warwick, UK*
[3]*Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, the Netherlands*

Place is a naturally diverse subject. It requires an interdisciplinary approach and must be addressed from different angles in order to fully reflect both the concept and specific places that shape people's everyday geographies. The diversity of available and required approaches to the study of places is well reflected in the first published articles in Transactions in GIS, which were recently published in response to a special collection in connection with the First International Symposium on Platial Information Science (Westerholt et al., 2018) last year. These articles include cartographic (Iosifescu Enescu et al., 2020), linguistic (Lai et al., 2020), socio-theoretical (Acedo and Johnson, 2020), and geo-demographic approaches (Ballatore and De Sabbata, 2020), and more articles are soon to follow. The contributions contained in this present volume for PLATIAL'19 also reflect a variety of fields and approaches of place research. These include linguistics, geography, semantics, psychology, politics, routing, and demography, truly filling this year's symposium motto with life!

An important aspect in connection with the concept of place is language and the way in which people express their mental models of lived spaces in linguistic expressions. Thora Tenbrink has delivered an inspiring and visionary keynote to the symposium, showing possible future ways to systematically explore the way people encode representations of place linguistically (Tenbrink, 2020). While much is known about spatial language, place is still a relatively unexplored field in terms of lingustic expressions of sensory-motor experiences and emotional attachments. The visions that Thora has shared with the audience point to possible ways of filling this gap, and have sparked many exciting discussions in Coventry last September.

Closely related to verbal statements of place are discursive practices, such as how people literally ask questions about places. With a view to possible future place-based GIS operations, Clare Davies presented a number of such questions and types of platial knowledge that people make use of in everyday life (Davies, 2020). Her research shows that most of these questions are not spatial and crisp (as in traditional GIS), but semantic and vague in nature. Clare also points out the importance of finding ways to formalize the questions listed in her paper in order to make a leap forward towards place-based GIS.

In order to talk about places and represent them meaningfully in the form of map or formal symbols, people must find a common ground – a shared understanding of what a place means. Maren Mayer, Daniel W Heck, and Franz-Benjamin Mocnik present a psychological contemplation of the complex negotiation processes connected with the search for shared mental models of places (Mayer et al., 2020). The paper uses the collaborative mapping project OpenStreetMap as an example to trace and visualize idiosyncratic concepts of places and their convergence to a common understanding of geographical entities. The presented research is an excellent contribution to a better understanding of shared, user-generated datasets.

R Westerholt and FB Mocnik (2020): *Introduction to the Second International Symposium on Platial Information Science*. In: FB Mocnik and R Westerholt (eds.), Proceedings of the 2nd International Symposium on Platial Information Science (PLATIAL'19), pp. 1–3

https://doi.org/10.5281/zenodo.3628847

PLATIAL'19 Second International Symposium on Platial Information Science (PLATIAL'19)
Coventry, UK; 5–6 September 2019

Place is not only of importance in GIScience and linguistics, but has been an outstanding topic in human geography for decades. Daniel Wagner, Rene Westerholt, and Alexander Zipf establish a link between GIScience and human geography by investigating how GIScience scholars make use of human-geographic place concepts (Wagner et al., 2020). The results reveal interesting patterns, including the frequent use of social media data in platial research and frequent references to Yi-Fu Tuan's work on place. These and other findings are an important impetus to gaining a better understanding of the theoretical foundations of research on place conducted in GIScience. The contribution presented here is thus a step towards a more informed debate on place.

Space and place are not independent concepts, but closely intertwined. Using Twitter data from the 2012 US Presidential Election and the 2014 Scottish Independence Referendum campaigns, Adrian Tear examines how the patterns of space and place in microblogging data vary (Tear, 2020). The results show that people geotagging with spatial coordinates are much less likely to use verbal toponymic references to named locations. Similarly, the same users link more non-platial extrinsic content in their tweets than non-geotagging users. These results indicate that the ways in which people conceptually grasp place and space are not independent, but at least interact in online communication. The presented results thus provide promising novel insights into how people use and represent places in digital tools like microblogging.

Social media are an important source of platial information that can be used for a number of applications. Madalina Gugulica and Dirk Burghardt present an approach to use the affective perceptions of urban environments extracted from social media to improve pedestrian routing applications (Gugulica and Burghardt, 2020). Typically, routing applications are designed to focus on optimizing travel time or distance. The approach presented here allows to add a human dimension to this rather technical view by also considering the way people react to geographical places. The innovative approach presented is a useful contribution to the future development of place-based human–machine interfaces and to the development of place-based technologies that will facilitate the daily life of many people.

The way we express ourselves verbally varies geographically. Justin van Dijk and Paul A Longley use this feature of local linguistic variation to trace the historical distribution of surnames in Great Britain (van Dijk and Longley, 2020). Using historical censuses and contemporary population registers, the study presented in this volume allows the characterization of places in terms of geodemographic characteristics. A total of 59,218 surnames have been studied and a hierarchy of places has been identified, with larger conurbations being more closely linked to other parts of Great Britain than smaller towns and rural areas. This provides an opportunity to introduce a further aspect of places into the current GIScience discourse and to link the topic to areas such as genealogy and the historical sciences.

Acknowledgements

We are grateful to everyone who contributed to making this symposium a big success! In particular, we want to thank Nicole Hengesbach, Jonathan Davies, and Alexander Noll (student helpers), as well as Sarah Doughty and Tracy Thompson (administrators) for their invaluable backstage support. We also feel very obliged to our keynote speakers Nigel Thrift and Thora Tenbrink for their very inspiring talks. Further, we are indebted to all members of the programme committee for their excellent reviews of the submissions to PLATIAL'19: Thomas Blaschke, Dirk Burghardt, Alexis Comber, Clare Davies, Ekaterina Egorova, Sara I Fabrikant, Michael F Goodchild, Krzysztof Janowicz, Karen Kemp, Grant McKenzie, Alenka Poplin, Simon Scheider, Kathleen Stewart, Thora Tenbrink, Nigel Thrift, Maria Vasardani, and Stephan Winter. Last but not least we are very grateful to all participants of the symposium who joined us at the University of Warwick in Coventry for making PLATIAL'19 a tremendous success!

ORCID

Rene Westerholt https://orcid.org/0000-0001-8228-3814
Franz-Benjamin Mocnik https://orcid.org/0000-0002-1759-6336

References

Acedo, Albert and Johnson, Peter A: *Home range and habitat: using platial characteristics to define urban areas from the bottom up*. Transactions in GIS, 2020. doi: 10.1111/tgis.12597

Ballatore, Andrea and De Sabbata, Stefano: *Los Angeles as a digital place: the geographies of user-generated content*. Transactions in GIS, 2020. doi: 10.1111/tgis.12600

Davies, Clare: *Place-based knowledge systems: human and machine*. In: Mocnik, Franz-Benjamin and Westerholt, René (eds.), *Proceedings of the 2nd International Symposium on Platial Information Science (PLATIAL'19)*. 2020, 33–41. doi: 10.5281/zenodo.3628865

van Dijk, Justin and Longley, Paul A: *Platial geo-temporal demographics using family names*. In: Mocnik, Franz-Benjamin and Westerholt, René (eds.), *Proceedings of the 2nd International Symposium on Platial Information Science (PLATIAL'19)*. 2020, 23–31. doi: 10.5281/zenodo.3628863

Gugulica, Madalina and Burghardt, Dirk: *Affective route planning based on information extracted from location-based social media*. In: Mocnik, Franz-Benjamin and Westerholt, René (eds.), *Proceedings of the 2nd International Symposium on Platial Information Science (PLATIAL'19)*. 2020, 51–60. doi: 10.5281/zenodo.3628877

Iosifescu Enescu, Cristina M; Bär, Hans-Rudolf; Beilstein, Matthias; and Hurni, Lorenz: *Place cookies and setting spiders in dream cartography*. Transactions in GIS, 2020. doi: 10.1111/tgis.12604

Lai, Juntao; Lansley, Guy; Haworth, James; and Cheng, Tao: *A name-led approach to profile urban places based on geotagged Twitter data*. Transactions in GIS, 2020. doi: 10.1111/tgis.12599

Mayer, Maren; Heck, Daniel W; and Mocnik, Franz-Benjamin: *Shared mental models as a psychological explanation for converging mental representations of place – the example of OpenStreetMap*. In: Mocnik, Franz-Benjamin and Westerholt, René (eds.), *Proceedings of the 2nd International Symposium on Platial Information Science (PLATIAL'19)*. 2020, 43–50. doi: 10.5281/zenodo.3628871

Tear, Adrian: *Geotagging matters? The interplay of space and place in politicized online social media networks*. In: Mocnik, Franz-Benjamin and Westerholt, René (eds.), *Proceedings of the 2nd International Symposium on Platial Information Science (PLATIAL'19)*. 2020, 61–72. doi: 10.5281/zenodo.3628879

Tenbrink, Thora: *The language of place: towards an agenda for linguistic platial cognition research*. In: Mocnik, Franz-Benjamin and Westerholt, René (eds.), *Proceedings of the 2nd International Symposium on Platial Information Science (PLATIAL'19)*. 2020, 5–12. doi: 10.5281/zenodo.3628849

Wagner, Daniel; Zipf, Alexander; and Westerholt, Rene: *Place in the GIScience community – an indicative and preliminary systematic literature review*. In: Mocnik, Franz-Benjamin and Westerholt, René (eds.), *Proceedings of the 2nd International Symposium on Platial Information Science (PLATIAL'19)*. 2020, 13–22. doi: 10.5281/zenodo.3628855

Westerholt, Rene; Mocnik, Franz-Benjamin; and Zipf, Alexander: *On the way to platial analysis: can geosocial media provide the necessary impetus? Proceedings of the First Workshop on Platial Analysis*. 2018. doi: 10.5281/zenodo.1475269

The Language of Place: Towards an Agenda for Linguistic Platial Cognition Research

Thora Tenbrink

School of Languages, Literatures and Linguistics, Bangor University, UK

Research on language in the interdisciplinary field of spatial cognition has identified multiple ways in which language represents mental representations of space, including object locations, spatial relationships, spatial problem solving such as wayfinding, and so on. Further, cognitive linguistic research reveals various ways in which language is based on physical experience. What remains under-explored is how the very fundamental human experience of place, in terms of its sensory and emotional attachments, is represented in language (other than works of art). Here I explore possible transfer avenues from linguistic spatial cognition to platial linguistic research.

Keywords: systematicity; methodologies; context; concepts; appreciation; platial language

History: received on 9 November 2019; published on 27 January 2020

1 Introduction

Spatial cognition research has benefited greatly from insights and methodologies in linguistics, adapted for the purpose of understanding how we think about space. How people *talk* about spatial locations and relationships systematically reflects how they *think* about space. This insight has led to a rich diversity of relevant interdisciplinary research and outcomes including GIS applications (Denis, 2017).

So far, pertinent research has focused far more on notions of *space* rather than *place*. The distinction between the two is symptomatic of the typical perspectives adopted in this research field, in spite of its inherent disciplinary diversity. Spatial notions tend to be abstract, formalizable, and context-free (as much as possible). How humans conceptualize space may be rooted in their experience, but the experience itself (in a *platial* sense) is rarely addressed in language-oriented spatial research. Instead, such research focuses on the linguistic expression of spatial perception and conceptions: where locations are, how they relate to each other, and how to get from A to B in familiar and unfamiliar environments. Route descriptions, for instance, have been investigated thoroughly in this tradition, showing how we understand and talk about space when we need to find our way. However, such research rarely captures what really matters for us in our environment, with some exceptions: for instance, pertinent insights concern the importance of landmarks for spatial processing, encompassing both visual salience and personal relevance (Caduff and Timpf, 2008).

Platial research, in contrast, directly addresses human experience, perception, and appreciation in a wider sense. Appreciating an environment does not only mean knowing where places (and landmarks) are, but – much more importantly – knowing what these places mean to us, how we relate to them, what makes them special. Conceptualizations of place must therefore be explored on the basis of how humans live in the world: which kinds of places play what kinds of roles; what are the boundaries of places in terms of their pragmatic reality and human experience; how and under what circumstances are

T Tenbrink (2020): *The Language of Place: Towards an Agenda for Linguistic Platial Cognition Research*. In: FB Mocnik and R Westerholt (eds.), Proceedings of the 2nd International Symposium on Platial Information Science (PLATIAL'19), pp. 5–12

https://doi.org/10.5281/zenodo.3628849

PLATIAL'19 Second International Symposium on Platial Information Science (PLATIAL'19)
Coventry, UK; 5–6 September 2019

emotional attachments developed for specific places; and so on. These perspectives, and their expression in language, may seem elusive and infinite, but the underlying principles are not. Many such principles may in fact be generic, as already demonstrated by pertinent platial research (Cresswell, 2014; Davies, 2018). To capture their linguistic expression, platial discourse needs to be addressed systematically, quite possibly along similar lines as traditional linguistic spatial cognition research. In the following, I will explore relevant research in linguistic spatial cognition with a view on transferability towards an envisioned agenda addressing notions of *place* through the lens of language.

2 Spatial Language and Cognition

Linguistic research in spatial cognition has a rich and varied tradition, albeit one that is no more than a few decades old. Linguistics is a relatively young academic field of its own, which started out with a focus on theoretical explorations of grammatical and lexical structures. In the past, there was thus little room for exploring the ways in which humans use specific domains of language to express specific domains of thinking. However, the surge in interdisciplinary spatial cognition research in the last part of the previous century has sparked a growing interest in the relationships between language and understanding of space. Here are some key insights.

2.1 Key Insights

Space is fundamental to human thinking for the simple reason that we grow up and live immersed in spatial environments (Newcombe and Huttenlocher, 2003; Plumert and Spencer, 2007). This obvious fact remains one of the major obstacles for establishing true artificial intelligence in computers and robots (Goswami and Vadakkepat, 2019): they do not benefit from the intense everyday experience of perceiving and freely moving around in space. Their strengths lie elsewhere, in the abstract computational procedures of their detached 'minds' that are not grounded in space (Beni, 2019). We, as humans, in contrast, understand life on the basis of its embedding in space. It matters to us where things are, how to get to places, how far away from us and from each other things and places are. These things concern our everyday lifes in multiple ways, and so we express them frequently in language. Everyday language therefore contains a host of information about space, reflecting the ways in which we perceive and conceive of space (Talmy, 2000).

This ubiquity has led to an effect that cognitive linguists explore in terms of transferred or 'metaphorical' usage: namely, that spatial terms are frequently found in abstract domains (Tyler and Evans, 2003). Take the previous sentence as an example, where neither *led to* nor *in* (which occurs twice in the sentence) retain their literal spatial meaning. In spite of the fact that the sentence does not seem particularly poetic, we can see the transfer of a path concept (a path is physically *leading to* somewhere) as well as a container concept (something is physically enclosed *in* something else) to a more abstract concept. Spatial terms with both literal and transferred senses are ubiquitous in language, aptly reflecting our physical everyday experience.

There are a number of principles about human spatial concepts that can be derived from the specific ways in which space is verbalized. For instance, we do not *think* (and rarely *speak*) in terms of metric or quantitative measures. Instead of 'The car is 22.35 cm away from the tree', we are much more likely to say 'The car is *near* (or *at*) the tree'. Talmy (1983) explored in detail how spatial language schematically structures space. In spite of the many possible relationships between objects in natural environments, linguistic terms (especially spatial prepositions such as *near* or *at*) draw on a very limited set of principles that can be detected by systematic analysis of linguistic resources and contextual usage. For instance, where exactly is *above*? The term immediately evokes the sense of a vertical relationship in our minds, but no specific distance, and not even a clear angle – it does not have to be anywhere specific. A *qualitative* notion of verticality is sufficient.

With the related term *over*, the context-dependent effects of functionality (Coventry et al., 1994) come into play. Whether or not an umbrella is *over* a person will often depend on where the rain comes from. Similarly, whether an object (say, a flower) is *in* another object (such as a vase) depends on *location control* rather than geometric containment (Coventry and Garrod, 2004). Clearly, the vase does not contain the flower, but it controls its location: if you move the vase, the flower will move with it. Notably, this specific effect is not part of the lexical entry for *in*. We intuitively know these things from our everyday experience, and apply them to our language use with ease.

Another fundamental insight into human spatial thinking comes from the analysis of a group of expressions commonly known as *projective* terms: *left*, *right*, *in front*, *behind*, and so on. The spatial relations expressed by these terms can be understood in various ways, depending on the underlying reference system and perspective (Levinson, 1996). For instance, *your left* is different from *my left* – and how large the *left* area is might differ according to the situation (Hayward and Tarr, 1995; Moratz and Tenbrink, 2006). Again, the use of a specific spatial term depends on how the speaker and hearer conceive of the spatial relationship in question.

Notably, in all of these cases, there are aspects of the environment that are not expressed in language. This relates to a very basic principle of communication: speakers express (and hearers understand) what is relevant for them, rather than aiming to somehow represent all details (whatever that might mean). Relevance is a major principle in communication (Sperber and Wilson, 1986), and it explains how people manage to fill in the many gaps in communication. For instance, no route description will ever be entirely unambiguous and complete (Tenbrink, 2012) but people will typically understand whatever they need in order to find their way to a destination. In these and other ways, spatial language (and its use in natural discourse) reflects the systematic principles according to which we understand the spatial world around us.

2.2 Theory: Language is Based on Experience

The insight that language is rooted in spatial experience has many implications, well beyond the fact that spatial language is ubiquitous in both literal and abstract senses (as outlined above). Famously, Lakoff and Johnson (1980) demonstrated how our physical experience affects our understanding of more abstract affairs, as reflected in metaphorical conceptions such as *good* is *up*: expressions such as *high spirits* and *feeling down* consistently represent positive experiences as higher up than negative ones, in line with spatial experiences such as having more of something good (like food or coins) frequently represents a higher pile, and somebody who feels good tends to be upright, while the sick may need to lie down.

More generally, the way a language develops represents the way its speakers experience the world – an insight already noted by Whorf (1956), who claimed that language represents the mass mind. Whorf's further speculations as to how the structures of a language, evolved on the basis of its speakers' thoughts, may in turn *constrain* its speakers' thoughts, led to a host of research investigating the intricate interdependencies between thought and language. More modern theories of embodied cognition (e.g., Wilson, 2002) focus primarily on the specific ways in which (and limits as to how) physical experience determines thoughts, and how exactly mental representations find their expression in language (Shapiro, 2019).

As part of this, contrary to previous assumptions concerning the autonomous status of language in the human mind (Chomsky, 1964), more recent theories explain how meanings are acquired through usage in context (Tomasello, 2009). This feat is substantially supported by fundamental cognitive mechanisms that help us bootstrap from known experience to novel insights, acquiring new meanings gradually on the basis of existing ones (Gentner, 2010). Taken together, these theories highlight the profound effects of experience on cognition and its representation in language.

3 The Experience of Place

The insights outlined so far leave little doubt that embodied experience is central to human life, as expressed in language. At a closer look, however, the vast majority of these insights concern what might be aptly characterized as *scientific* experience, on a personal basis. As we discover the world and its mechanisms, we develop concepts and linguistic terms to express them, in line with conventions in the society we live in. This main principle drives the specific ways in which language evolves and thoughts are expressed – generally within and across cultures, and individually in specific situations.

Traditionally, *cognition* is understood as separate from *emotion*, leading to distinct areas of research (a tradition that has been questioned; Pessoa, 2008). The linguistic principles discussed so far are firmly situated in the realm of cognition, and consistently leave aside any aspects that concern *appreciating* rather than *understanding*. In contrast, notions of *place* fundamentally concern human emotional connections with spatial locations that go beyond personal scientific insight. If human thought is as deeply intertwined with spatial experience as is now widely understood, it is time to account for the linguistic expression of all aspects of this experience – including the non-scientific, emotional experience of appreciating places. The following sections will explore existing insights in this regard and then point to what might be aptly recognized as a gap.

3.1 Pertinent Research Areas and Insights

In spite of the fact that spatial cognition research focuses on scientific rather than affective aspects of experiencing space, a range of insights emerging from the field do pertain to a more personal level, highlighting certain aspects of platial appreciation. All experiences of space are personal, perceived individually in time and space. This is aptly represented by the system of *deixis* in language, which captures the here-and-now of our experience (Fillmore, 1982) using personal pronouns, distinctions between *here* and *there*, or *now* and *then*. Recent research demonstrates that the use of deictic terms may be influenced by factors such as distance, ownership, visibility, and familiarity (Coventry et al., 2014) – aspects that are far more personal than the scientific experience of space.

Further pertinent insights include the central role of *relevance* for cognition as well as communication (Sperber and Wilson, 1986), as well as notions of *salience* that guide focus of attention in various ways (Chiarcos et al., 2011). Both relevance and salience are rooted in personal experience. Objects and places (and abstract ideas) are often relevant to us for a personal reason related to a specific context – this may well be charged with emotional associations. Similarly, objects (or landmarks) may stand out as particularly salient for us not only because of visual contrast to other objects, but also because of their special meaning for us (Caduff and Timpf, 2008). Also, it is well-known in wayfinding research that although route choices can be predicted to some extent by generic heuristics (Hochmair and Frank, 2002), there will always be a certain amount of individual variation based on people's preferences (Hölscher et al., 2011).

In platial research, it is more widely accepted that knowledge of place is semantic, i.e., meaningful, and based on emotional significance (Davies, 2018). This insight has motivated much research in human geography and geographic information science, and some systematic studies of language (Purves and Derungs, 2015; Stock, 2008; Winter and Freksa, 2012). In this realm, much research targets the exploration of *sense of place*, including effects of social and geographical contexts on how places are perceived (Hay, 1998), as well as beliefs about the relationship between oneself and a place, in terms of ownership and behavioural commitments (Jorgensen and Stedman, 2001), or notions of rootedness, belonging, and routines (Buttimer and Seamon, 1980).

Insights such as these complement extensive research in the humanities that has long recognized the significance of how language (and in particular the language of art) represents affective aspects of our spatial experience. Multiple works of poetry, literature, music, and other art forms vigorously express experience and emotion, beyond personal scientific insights – and a host of academic research is available to discuss this in much depth. Take, for instance, John Denver's 'Take Me Home, Country Roads', released 1971 – a song that became popular around the world well beyond audiences who are themselves at home in West Virginia (as described in the song's lyrics), supported by the vivid visual imagery that successfully conjures up emotions of 'almost heaven' (Byklum, 1994).

Notions of *home* as a place with particular emotional significance are also frequently debated in works of fiction and their academic treatment, recognizing the effects of such a spatial location on human experience well beyond its existence as a physical or geographical entity (Rubenstein, 2001). Notably, even though *home* is a very personal notion, the described effect is very generic and very describable (and hence explored widely in the humanities), far from having to be discarded as 'unknowable' along with other personal experiences that are too diverse and individual to be captured systematically.

3.2 The Gap

The overview in the previous section suggests that most insights in platial research, so far, have been generated as a by-product of systematic research on spatial language and cognition, by extensive targeted research in areas of human geography or environmental psychology, and by humanities research that focuses on the affective significance of places in human lifes as represented across many art forms. While significant insights have been generated in all of these areas, language as a representation of human experience could be exploited far more directly and extensively.

Systematic research could target the linguistic expression of *platial* notions, as opposed to primarily *spatial* ones, doing justice to the fact that emotional and spatial experience are deeply intertwined and ultimately inseparable. It seems that even though this insight has long been recognized in other academic realms, it has not quite reached linguistic spatial cognition research – quite possibly hampered by the traditional separation of cognition from emotion. As a consequence, any relevant insights that may be generated alongside those pertinent to cognition proper tend to be marginalized, rather than being taken seriously as manifestations of basic human experience.

While in-depth qualitative insights based on (particularly literary) language samples have been generated extensively, more generic or quantitative approaches known from linguistic spatial cognition research could be transferred to explore platial notions in a range of ways. Moving on from the wide-ranging recognition that language is based on experience, theoretical approaches (in cognitive linguistics or elsewhere) could specifically target the systematic features of language that express platial notions. Computational approaches need to be enhanced to implement platial language and concepts more thoroughly, accounting for the generic effects that platial experience has on human thinking. To inform both theory and computation, empirical studies are required that systematically address platial language use across various generic types of scenarios. This would parallel the host of research targeting spatial language use in many different contexts, which has successfully identified generic principles that affect spatial language use and highlight important facets of human spatial cognition.

4 Investigating Place Through Language: Towards an Agenda

Language is, as we have seen, an excellent medium for investigating human concepts systematically. Humans have a desire to communicate their thoughts and feelings to others, and thus represent both scientific and emotional experiences in language. Moreover, linguistics is often seen as situated between the science–humanities divide (Pulgram, 1969) and is therefore an obvious candidate discipline for addressing a concern that appears to cross over this divide.

Above all, how humans live in the world, and thus experience it, is key. Personal experience leads not only to generic scientific insights but equally to the development of a *sense of place* related to various locations in the environment. Although specific sense of place experiences are certainly unique to each individual in relation to their platial attachments, we all use the existing repertory of our language to represent our thoughts and feelings – and language thus follows rules and principles that will allow for capturing generic structures that affect our use of language.

A systematic agenda for the investigation of place through linguistic analysis could therefore target questions such as the following, mapping the existing analysis principles for *spatial* language to the investigation of *platial* language:

- What are the principles according to which places are conceptualized and verbalized?
- Which kinds of places play what kinds of roles in human discourse?
- What are the boundaries of places in terms of their pragmatic reality in everyday language use?
- How and under what circumstances are emotional attachments expressed for specific places?
- What are the main categories and concepts represented in language in platial contexts?
- What are the overarching principles behind these categories?
- What do they reveal about the mental representation of place?
- How is this linguistic-conceptual repertory used in discourse, across different settings?

To give a sense of how this might work, consider the following statement expressing platial notions of a local who describes an area in North Wales:

> *I like the old bridge and looking down at the water beneath, the way the tide changes and reveals and hides different islands. The cormorants nesting on the bridge and the turbulence of the currents are fascinating to watch.*

The description clearly reflects both spatial and emotional aspects, closely tied together in the way the place is experienced. Spatial aspects are represented by much-researched prepositions such as *down*, *at*, *beneath* and *on*. In contrast, the platial language is far more diverse and striking in this short text. There are verbs of volitional perception (*looking down* and *to watch*), reflecting what the observer chooses to perceive in their spatial environment, and conveying a sense of connection to nature, to the water and the tide, the cormorants and the turbulence of currents. There are direct terms of appraisal (*like*, *fascinating*) that highlight specific aspects of the place. And there are poetic features that convey a sense of admiration, contrasting with what one might expect in a neutral location description (*beneath* rather than *below*; the rhythmic parallelism in *changes and reveals and hides*).

Linguistic features of this kind may seem related to individual style, but already there is ample indication (in abundant humanities research on works of art, and in existing data) that what we see here goes beyond the random example of an enthusiastic lover of nature. The observations we have just noted here may well indicate some of the generic ways in which speakers express platial notions.

5 Conclusion

Notions of place are central to human experience, and they are pervasive in discourse. As our everyday lives happen in spatial environments, we frequently speak about locations and places, often combining 'scientific' and 'emotional' representations. In academic research, these have been treated separately to a large extent, leaving a gap that could be addressed by a targeted analysis of language, combining insights and methods across research fields. This will allow for more systematic and generic insights, and ultimatively a new understanding of what it means for language to be based on experience – integrating factual and affective facets.

Acknowledgements

I would like to thank the organizers of PLATIAL'19 for inviting me, the workshop participants for highly inspiring contributions and discussions, and Clare Davies for invaluable comments on a previous version of this paper.

ORCID

Thora Tenbrink https://orcid.org/0000-0002-7986-1254

References

Beni, Majid D: *On the origin of mental representations*. Biosystems, 184, 2019, 103995. doi: 10.1016/j.biosystems.2019.103995

Buttimer, Anne and Seamon, David: *The human experience of space and place*. New York, NY, USA: Taylor and Francis, 1980. doi: 10.4324/9781315684192

Byklum, Daryl: *Geography and music: making the connection*. Journal of Geography, 93(6), 1994, 274–278. doi: 10.1080/00221349408979833

Caduff, David and Timpf, Sabine: *On the assessment of landmark salience for human navigation*. Cognitive Processing, 9(4), 2008, 249–267. doi: 10.1007/s10339-007-0199-2

Chiarcos, Christian; Claus, Berry; and Grabski, Michael: *Salience: multidisciplinary perspectives on its function in discourse*. Berlin: de Gruyter, 2011

Chomsky, Noam: *Current issues in linguistic theory*. The Hague: de Gruyter, 1964

Coventry, Kenny R; Carmichael, Richard; and Garrod, Simon C: *Spatial prepositions, object-specific function and task requirements*. Journal of Semantics, 11, 1994, 289–309. doi: 10.1093/jos/11.4.289

Coventry, Kenny R and Garrod, Simon C: *Saying, seeing and acting. The psychological semantics of spatial prepositions*. London, UK: Psychology Press, 2004. doi: 10.4324/9780203641521

Coventry, Kenny R; Griffiths, Debra; and Hamilton, Colin J: *Spatial demonstratives and perceptual space: describing and remembering object location*. Cognitive Psychology, 69, 2014, 46–70. doi: 10.1016/j.cogpsych.2013.12.001

Cresswell, Tim: *Place. An introduction*. Hoboken, NJ, USA: Wiley and Sons, 2014

Davies, Clare: *Place and placing locations: a cognitive perspective*. In: Westerholt, Rene; Mocnik, Franz-Benjamin; and Zipf, Alexander (eds.), *Proceedings of the 1st Workshop on Platial Analysis (PLATIAL'18)*. 2018, 15–20. doi: 10.5281/zenodo.1472737

Denis, Michel: *Space and spatial cognition: a multidisciplinary perspective*. London, UK: Taylor and Francis, 2017. doi: 10.4324/9781315103808

Fillmore, Charles J: *Towards a descriptive framework for spatial deixis*. In: Jarvella, Robert J and Klein, Wolfgang (eds.), *Syntax speech, place and action: studies in deixis and related topics*, Chichester, UK: Wiley, 1982. 31–59

Gentner, Dedre: *Bootstrapping the mind: analogical processes and symbol systems*. Cognitive Science, 34(5), 2010, 752–775. doi: 10.1111/j.1551-6709.2010.01114.x

Goswami, Ambarish and Vadakkepat, Prahlad: *Humanoid robotics: a reference*. Dordrecht: Springer, 2019. doi: 10.1007/978-94-007-6046-2

Hay, Robert: *Sense of place in developmental context*. Journal of Environmental Psychology, 18(1), 1998, 5–29. doi: 10.1006/jevp.1997.0060

Hayward, William G and Tarr, Michael J: *Spatial language and spatial representation*. Cognition, 55(1), 1995, 39–84. doi: 10.1016/0010-0277(94)00643-Y

Hochmair, Hartwig H and Frank, Andrew U: *Influence of estimation errors on wayfinding-decisions in unknown street networks – analyzing the least-angle strategy*. Spatial Cognition and Computation, 2(4), 2002, 283–313. doi: 10.1023/A:1015566423907

Hölscher, Christoph; Tenbrink, Thora; and Wiener, Jan M: *Would you follow your own route description? Cognitive strategies in urban route planning*. Cognition, 121(2), 2011, 228–247. doi: 10.1016/j.cognition.2011.06.005

Jorgensen, Bradley S and Stedman, Richard C: *Sense of place as an attitude: lakeshore owners attitudes toward their properties*. Journal of Environmental Psychology, 21(3), 2001, 233–248. doi: 10.1006/jevp.2001.0226

Lakoff, George and Johnson, Mark: *Metaphors we live by*. Chicago, IL, USA: The University of Chicago Press, 1980

Levinson, Stephen C: *Frames of reference and Molyneux's question: crosslinguistic evidence*. In: Bloom, Paul; Garret, Merril F; Nadel, Lynn; and Peterson, Mary A (eds.), *Language and space*, Cambridge, MA, USA: MIT Press, 1996. 109–169

Moratz, Reinhard and Tenbrink, Thora: *Spatial reference in linguistic human-robot interaction: iterative, empirically supported development of a model of projective relations*. Spatial Cognition and Computation, 6(1), 2006, 63–107

Newcombe, Nora S and Huttenlocher, Janellen: *Making space: the development of spatial representation and reasoning*. Cambridge, MA, USA: MIT Press, 2003

Pessoa, Luiz: *On the relationship between emotion and cognition*. Nature Reviews Neuroscience, 9(2), 2008, 148–158. doi: 10.1038/nrn2317

Plumert, Jodie M and Spencer, John P: *The emerging spatial mind*. Oxford, UK: Oxford University Press, 2007. doi: 10.1093/acprof:oso/9780195189223.001.0001

Pulgram, Ernst: *Sciences, humanities, and the place of linguistics*. Linguistics, 7(53), 1969, 70–92. doi: 10.1515/ling.1969.7.53.70

Purves, Ross S and Derungs, Curdin: *From space to place: place-based explorations of text*. International Journal of Humanities and Arts Computing, 9(1), 2015, 74–94. doi: 10.3366/ijhac.2015.0139

Rubenstein, Roberta: *Home matters*. New York, NY, USA: Palgrave Macmillan, 2001. doi: 10.1057/9780312299750

Shapiro, Lawrence: *Embodied cognition*. London, UK: Taylor and Francis, 2019, 2nd edn. doi: 10.4324/9781315180380

Sperber, Dan and Wilson, Deirdre: *Relevance: communication and cognition*. Oxford, UK: Blackwell, 1986, 2nd edn.

Stock, Kristin: *Determining semantic similarity of behaviour using natural semantic metalanguage to match user objectives to available web services*. Transactions in GIS, 12(6), 2008, 733–755. doi: 10.1111/j.1467-9671.2008.01128.x

Talmy, Leonard: *How language structures space*. In: Pick, Herbert L and Acredolo, Linda P (eds.), *Spatial orientation: theory, research, and application*, Boston, MA, USA: Springer, 1983. 225–282. doi: 10.1007/978-1-4615-9325-6_11

—— *Towards a cognitive semantics*. Cambridge, MA, USA: MIT Press, 2000

Tenbrink, Thora: *Relevance in spatial navigation and communication*. Proceedings of the International Conference on Spatial Cognition, 2012, 358–377. doi: 10.1007/978-3-642-32732-2_23

Tomasello, Michael: *Constructing a language*. Cambridge, MA, USA: Harvard University Press, 2009

Tyler, Andrea and Evans, Vyvyan: *The semantics of English prepositions: spatial scenes, embodied meaning and cognition*. Cambridge, UK: Cambridge University Press, 2003

Whorf, Benjamin: *Language, thought, and reality: selected writings*. Cambridge, MA, USA: MIT Press, 1956

Wilson, Margaret: *Six views of embodied cognition*. Psychonomic Bulletin and Review, 9(4), 2002, 625–636. doi: 10.3758/BF03196322

Winter, Stephan and Freksa, Christian: *Approaching the notion of place by contrast*. Journal of Spatial Information Science, 5, 2012, 31–50. doi: 10.5311/JOSIS.2012.5.90

Place in the GIScience Community – an Indicative and Preliminary Systematic Literature Review

Daniel Wagner[1], Alexander Zipf[1], and Rene Westerholt[2,1]

[1]*Institute of Geography, Heidelberg University, Germany*
[2]*Centre for Interdisciplinary Methodologies, University of Warwick, UK*

The concept of place has recently gained importance in geographical information science (GIScience). One reason for this is the emergence of user-generated geographic information, which partially represents subjective everyday geographical encounters. No consensus, however, on how to deal with place in GIScience has yet been reached. This paper presents a systematic literature review providing an overview of how parts of the GIScience community currently use the concept of place as it is understood in human geography. The results suggest that most place related GIScience scholars refer to the humanistic tradition of geography focusing on the essence of experiences of place. Further, it is found that geotagged data published online are a major driver of place-based research, whereas scientific data (e.g., surveys) are less commonly found in respective papers. Many researchers make use of exploratory approaches, which may reflect the early stage at which place-based GIScience research still sits. We also identify a difference between the approach core members of GIScience take and those working on the edge of the field. Thereby, the former often work more conceptually than the latter. The results of this preliminary review inform the current GIScience discourse on place by important evidence about the intellectual standpoints of GIScience scholars, thus fostering future research into place.

Keywords: place; GIScience; systematic literature review

History: received on 12 July 2019; accepted on 16 August 2019; published on 27 January 2020

1 Introduction

The use of user-generated information, partially depicting the world from the everyday perspective of normal people, has become commonplace in geographical information science (GIScience). Notions of place that are geared towards experienced and perceived space (opposed to the geometric notion that is usually used) are therefore currently gaining popularity (Purves et al., 2019; Westerholt et al., 2018a). The concept of place, however, is fuzzy and hard to grasp. Further, various academic disciplines developed their own subject-related definitions and vocabulary concerning place, oftentimes through the lenses of time and philosophical currents. A consistent GIScience understanding of place does not yet exist (Merschdorf and Blaschke, 2018). The vision of a place-based GIS (Goodchild, 2011), however, requires an unambiguous definition of place, its formalization, and ways to extract meaningful information from subjective user-generated data (Merschdorf and Blaschke, 2018).

Several publications appeared using various concepts of place or place-based data (e.g., Chen et al., 2018; Gao et al., 2017; Scheider and Purves, 2013; Winter and Freksa, 2012). Many place-related publications in GIScience thereby make use of conceptual frameworks borrowed from human geography and apply these in different contexts. This conceptual variety, indicating a lack of consensus, motivates

D Wagner, A Zipf, and R Westerholt (2020): *Place in the GIScience Community – an Indicative and Preliminary Systematic Literature Review.* In: FB Mocnik and R Westerholt (eds.), Proceedings of the 2nd International Symposium on Platial Information Science (PLATIAL'19), pp. 13–22

https://doi.org/10.5281/zenodo.3628855

PLATIAL'19 Second International Symposium on Platial Information Science (PLATIAL'19)
Coventry, UK; 5–6 September 2019

the main research question of the present contribution: *In which ways do GIScience and cognate scholars make use of the concept of place?*

We address this research question through a systematic literature review. In order to answer the research question on the basis of the identified corpus, all the collected articles are disassembled with regard to a range of aspects. That is, we extract the ways in which the authors make intellectual use of the concept of place in their contributions. This allows us to draw conclusions about the GIScience scholars' understanding of the concept of place and how they use it. One way we investigate this is to look at the referenced geographical literature. Another path taken is to investigate the methodological approaches applied, as these may hint on the understanding of place authors make use of. Another indication we consider is the identification of research objectives found in the records. The results obtained show clear indications of how the community handles place-based data. They also suggest potential future directions towards developing an unambiguous GIScience definition and understanding of the concept of place.

The remainder starts out by briefly introducing the concept of place in Section 2. Section 3 then outlines the approach taken for conducting the systematic literature review. The results of this review are presented in Section 4, before they are discussed and concluded in Section 5.

2 Place

Place has been of central interest to philosophers and geographers alike. The terms *place* and *platial*[1] have been used in different ways and contexts – from Aristotle (Drum, 2011) over Yi-Fu Tuan (Tuan, 1977) to Mike Goodchild (Goodchild, 2011). The ancient Greek philosophers Plato and Aristotle were probably the first to formulate a systematic philosophy of place. They coined the terms *topos* and *chôra* resembling the contemporary geographic concepts of *space* and *place* (Agnew, 2011; Casey, 1997). They thus distinguished two meanings of geographical space with space (topos) referring to a void location without qualities and place (chôra) being considered space imbued with meaning and identity: 'Place [. . .] is a part of the terrestrial surface that is not equivalent to any other, that cannot be exchanged with any other without everything changing. Instead, with space, each part can be substituted for another without anything being altered' (Farinelli, 2003, p. 11). This dualism thus has a long tradition and is still found in the contemporary literature.

Contemporary debates about place (including those in GIScience) are strongly influenced by the human-geographic discourse that began in the 1970s. Three general branches are typically distinguished: regional-geographic accounts of place, ideas from humanistic geography, and the process-oriented viewpoints of radical geographers (Cresswell, 2014). One understanding that has emerged since then is that the crisp dualism between space and place cannot hold true. Yi-Fu Tuan, one of the most influential scholars writing on place, noted in 1977 that '[w]hat begins as undifferentiated space becomes place as we get to know it better and endow it with value' (Tuan, 1977, p. 6). This makes clear that there is an inherent link between the abstract notion of space and place in the sense of lived space (Soja, 2008). GIScience scholars are facing the challenge to synthesize this complex concept into a notion that is available to formalization in a more technical context.

Any mathematical and computer-based approach to place requires formalized input, standardized and defined rules, and well-defined concepts and terminology (Goodchild, 2011). Another component of place is its dependence on context. People from different backgrounds have different ways to experience their everyday geographies. This is, for instance, reflected in the various ways in which the concept of place is expressed (and expressible) in different languages (Blaschke et al., 2018). Still, despite place being a heterogeneous and fuzzy concept, there also exists a core to it, a shared understanding that all languages treat space more as a container while place is usually interwoven with notions of human experience and perception. In order to work effectively and to ensure the comparability of research approaches, scholars should thus focus on this common core, which may serve as the basis for a thorough future understanding of place in GIScience.

3 Methodological Approach

The approach taken to answer the research question is a systematic literature review. We thereby borrow elements from Borrego's guidelines for systematic reviews in developing interdisciplinary fields (Borrego et al., 2014). In addition, we have classified the authors into two groups: core members of the GIScience community (evidenced by clear indications from their articles) and contributors from the fringes of the field. The way we have classified authors is as follows: Authors stating explicitly to be GIScientists in their papers are considered core members. As this may be too strict, we have relaxed the condition by also considering those core who explicitly mention the GIScience community and work on a topic from the core research agenda as proposed by Mike Goodchild and revisited by Thomas Blaschke and Helena Merschdorf (Blaschke and Merschdorf, 2014; Goodchild, 1992, 2010). The fringe of GIScience as utilized here then comprises authors who address at least two topics from that agenda alongside place. We anticipate that this is just one possible way to break down the community into finer parts. Still, it allows to gain a clearer picture of how the community approaches the topic of place. Our framework consists of four main steps (see Figure 1), which are outlined in the following subsections.

3.1 Inclusion and Exclusion Criteria

The protocol presented is the outcome of an iterative refinement procedure, which has improved and supplemented the initial version to improve the quality of the study. The main selection criteria are as follows: We only consider records written in English, as most publications within the GIScience community are published in this language. Most records are extracted from two main databases: *Thompson Reuters Web of Science* (multidisciplinary) and *ACM Digital Library* (focus on computer science). These cover a broad range of the relevant literature. In addition, we have manually added the proceedings of the PLATIAL'18 workshop published on ZENODO (Westerholt et al., 2018b), as well as the proceedings of the tenth International Conference on Geographic Information Science (GIScience 2018), since these contain relevant recent records. The temporal interval is set to [1991, 2019], given that the GIScience community started operating under its name in 1991 with Mike Goodchild coining the term at a specialist meeting in that year (Goodchild, 1992, 2010). Clearly, this is a limitation as relevant literature from before 1990 might be available, too. Yet, setting 1990 as our start date reduces the variance and noise by constraining the considered community. Given the technical restrictions of our approach in terms of abstract screening (see below), only records with an abstract can be taken into consideration. It would otherwise not be possible to stick to the framework developed. This strategy has certainly led to the exclusion of relevant book chapters and other types of manuscripts. Subsequent work should take this into account. Finally, all records eligible need to discuss the concept of place itself (at a conceptual level), methodological approaches towards place, or the application of the concept within the GIScience community. Table 1 provides an overview of all queried databases and the search strings used.

3.2 Record Identification

The inclusion criteria led to the retrieval of 2,140 academic records from the Web of Science database (referred to as WOS hereafter) and 149 records from the ACM Digital Library (referred to as ACM hereafter). Removing duplicates ($n = 137$), records written in languages other than English ($n = 2$), records published before 1990 ($n = 1$), and those lacking an abstract ($n = 688$) has lead to a final inclusion of 1,461 academic records fulfilling all inclusion criteria. The corpus considered is structured as follows: 63% journal articles, 3% conference contributions, 18% book reviews, 7% editorials, 5% specialist meeting abstracts, 3% reviews, and 1% letter. This composition explains the high number of records lacking abstracts, as some of these categories do not normally come with abstracts (e.g., specialist meeting abstracts). The 1,461 records identified this way have entered the screening phase of the study.

3.3 Screening

The first screening step is the semi-automated abstract screening of the remaining 1,461 academic records. This is performed to remove unsuitable records lacking connection to the concept of place, to the GIScience community, or to the research question. To achieve this, we have compiled three

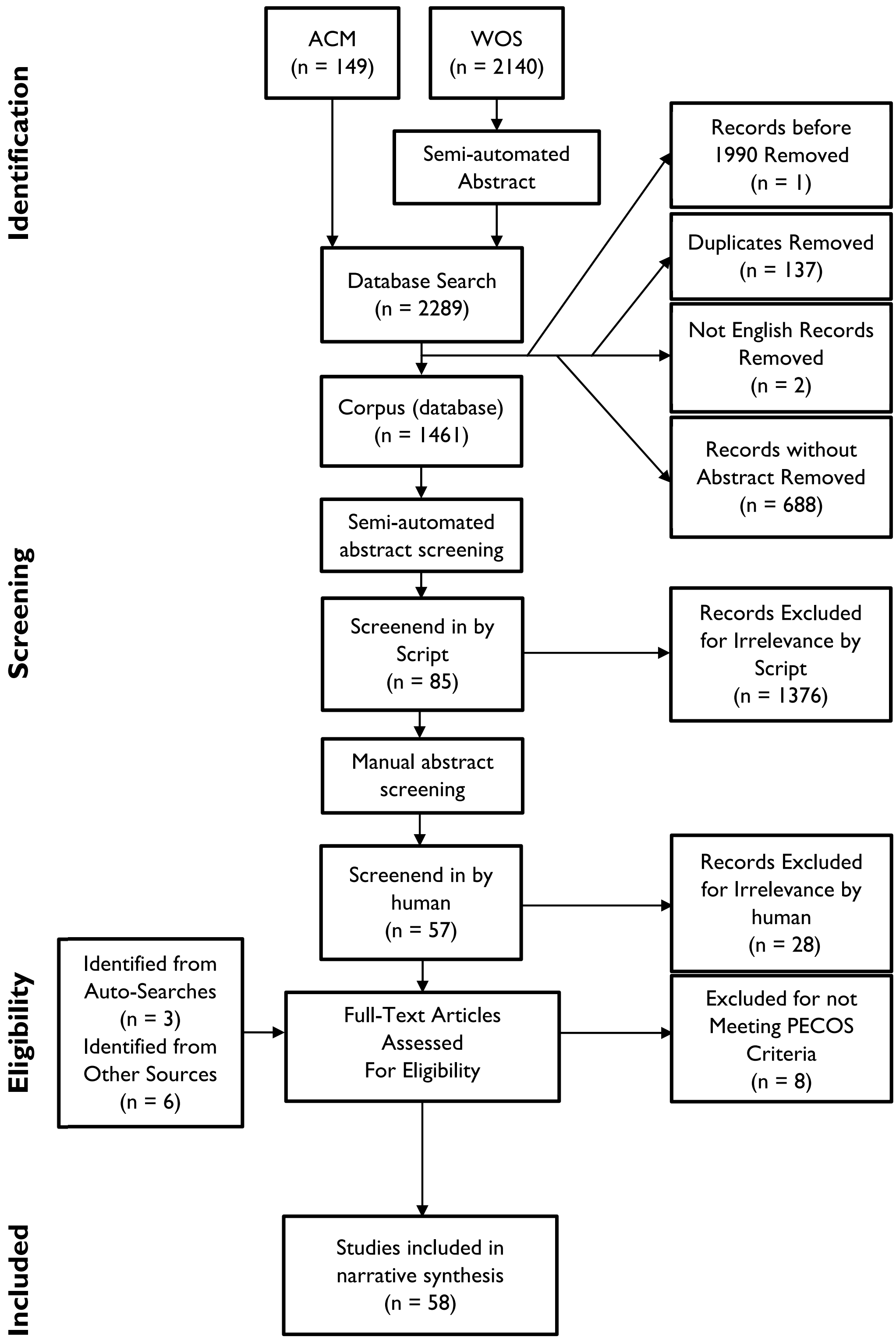


Figure 1: Systematic literature review approach. Overview of all stages of the systematic review.

Table 1: Database search strategy. The databases and search query strings used to carry out the review.

Database	Search Query String
Web of Science	`(TI=("place" OR "places" OR "platial" OR "place based") AND TI=(GIScience OR Geoinformatics OR GIS OR "Geographic Information Science" OR "Geographic Information" OR ppgis OR "public participation gis" OR "spatial cognition" OR "cognitive reference" OR representation OR define OR defines OR defining OR defined OR definition OR definitions OR concept OR concepts OR formalization OR formalize OR formalized OR formalizing OR space OR spaces OR spatial OR methodology OR method OR technique OR application OR utilization)) AND LANGUAGE: (English), SSCI Timespan=1991-2019`
ACM Digital Library	`+acmdlTitle:(place platial) +acmdlTitle:(GIScience Geoinformatics GIS "Geographic Information Science" "Geographic Information" ppgis "public participation gis" "spatial cognition" "cognitive reference" representation define defines defining defined definition definitions concept concepts formalization formalize formalized formalizing space spaces spatial methodology method technique application utilization)`

bags of keywords: one related to the concept of place, another one representing GIScience, and a third one reflecting the research question. Only records containing keywords from all three bags are taken into account further. In addition, a position-based syntactical context analysis was applied to exclude records that do not contain place and space-related terms in mutual vicinity of 10 positions. The choice of a maximum distance of 10 positions is to a certain extent arbitrary and should be replaced in further research by an improved syntactic procedure. Still, our results indicate that even the straight-forward approach taken here allows to get rid of general discussions about geography, where these terms would appear naturally. Performing this step leaves 85 papers for further manual abstract screening. During the manual screening step, another 28 records were excluded, as these did not discuss the concept of place (e.g., 8 publications were discussing hippocampal place cells), or used the term place in unintended ways (e.g., as a synonym for space, which was the case with more than 10 publications from computer science). It also turned out that some publications were not related to the GIScience community though they were discussing place as intended (more than 10 publications). After this step, 57 records were left for the final eligibility and quality assessment.

3.4 Quality Assessment

The final step is based on assessing the full text of the 57 records left from the screening step, 8 of which failed to match the criteria of this study as identified by manual screening. In all, 58 records fulfilled all criteria and were thus considered eligible. The quality assessment tasks (QA) defined in regard to the research question help to evaluate certain individual aspects:

- QA 1: *Is the concept of place utilized articulated clear enough to allow any assessment?* This criterion determines whether the explanation is adequate to fully identify the concept used.
- QA 2: *Do the authors modify a concept of place or do they develop a novel conceptual contribution?* This criterion focuses on the originality of an author's contribution and engagement with place.
- QA 3: *Do the authors develop new methods regarding the concept of place?* This criterion shows the depth of the authors' engagement with the concept of place at a methodological level.
- QA 4: *Is the application of the concept of place original and creative in an empirical sense?* If so, this shows that the authors are seriously engaging in empirical investigations of particular places.

- QA 5: *Are the limitations of the concept of place used clearly outlined?* This demonstrates a strong awareness for, and a critical reflection of, potential weak points of the approach applied.

Points were awarded to rank the individual records. Thereby, 1 point was awarded if a criterion was fully met, 0.5 points indicate partial fulfilment, and 0 points were given when a record failed to meet a criterion. Based on this, the records were classified into the categories 'good' (5–3.5 points), 'fair' (3.0–2.5), and 'poor' (< 2.5 points), which has identified 30 studies to be of good quality, 26 were found to be fair, whereas only 2 studies were considered poor.

3.5 Limitations

The approach taken has several limitations. The number of databases searched is two and therefore small. This is justified by the broad scope of the databases and because the scope of this study is relatively narrow. However, there is a chance that some relevant records might have been missed out. In addition, only English-language publications are considered, which leads to an absence of publications in other languages. However, we regard this as only a minor limitation, as the majority of the scholarly literature on the topic in question is written in English. Further, the search terms should be extended in future research to better match some areas that may be underrepresented in this study. For instance, keywords like 'spatial cognition' are contained, but other terms representing psychological aspects more comprehensively may be required in order to better reflect the importance of that field (see Davies, 2018). Further research related to identifying relevant place-related keywords could optimize the results further. The manual screening has been conducted using a detailed protocol but was not verified using a peer-review including a larger number of scholars. A last point worth mentioning is that our decision for a proximity-based context analysis might have given preference to records referring to the works of humanistic geographers. Nevertheless, the results discussed in the next section are still meaningful in the light of our research question.

4 Results

Among the retrieved records are 48 journal articles, 4 short papers, 4 conference proceeding papers, and 2 literature reviews. All those records were published between 1994 and 2019, with 39 of them being published after 2009 and at increasing annual publication rates. This finding supports the observation that the concept of place is becoming more popular recently within the GIScience community (Purves et al., 2019; Westerholt et al., 2018a). Institution-wise, the authors mainly come from the fields of geography (ca. 50%) and GIScience/geoinformation (ca. 30%), as well as other fields like architecture, cartography, remote sensing and, history (ca. 20% all together). A closer look at the authors reveals that Thomas Blaschke participated in three publications. Tim Cole, Song Gao, Alberto Giordano, Helena Merschdorf, and Emmanuel Papadakis have all participated in two publications. These are the most active authors identified in our corpus. Nevertheless, some key authors like Ross Purves are absent in the results. This indicates that some important keywords may be missing in our approach presented. The results presented in this short paper can thus be considered an impetus to the conception of a full assessment of GIScience's involvement with place.

4.1 Geographic Place Concepts Applied

We found that within the 58 retrieved academic records 20 different concepts of place were applied. Still, 13 records did not refer to any particular concept of place from the literature. Out of the 20 different concepts, 13 are used less than 3 times and 9 of them were only used once. The core members of the GIScience community appear to use a greater variation of concepts. Among them, more than 57% of the concepts used appeared once while for the records assigned to the contributors this holds true for 47%. The concepts of place that were applied most often are those from the phenomenological understanding of place proposed by Yi-Fu Tuan and related scholars. Tuan's concept (Tuan, 1977) was applied 14 times (members: $n = 6$; contributors: $n = 8$), closely followed by the work of Agnew (1987), which was used 6 times (members: $n = 3$; contributors: $n = 3$). Cresswell's concept (Cresswell, 1996) was applied 4 times (members: $n = 1$; contributors: $n = 3$), and Curry's concept (Curry, 1999) also was applied 4 times (members: $n = 2$; contributors: $n = 0$). Finally, Lefebvre's related concept of social space (Lefebvre and Nicholson-Smith, 1992) was applied twice (members: $n = 0$; contributors: $n = 2$).

4.2 Types of Place-Based Data Used

We have also briefly looked into the types of place-based data used across the corpus identified. The most frequently used type of data is online geotagged data (like those extracted from social media; contributors: 22%; members: 21%), followed by interview data acquired for the purpose of investigating specific places (contributors: 24%; members: none), secondary data identified from the academic literature (contributors: 11%; members: 43%), and traditional GIS data (e.g., administrative data sets; contributors: 9%; members: 29%).

4.3 Methodologies Applied

We identified 11 different methodological approaches within the retrieved records. The most prevalent type of methodology was case studies being found in roughly more than 36% of all records (contributors: 36%; members: 38%). Another frequently found approach is data exploration with a coverage of 19% (contributors: 21%; members: 8%), followed by literature review and analysis with an approximate share of 12% (contributors: 7%; members: 31%). Further popular approaches found are methodological frameworks to investigate place-related data (contributors: 7%; members: 23%), and social-scientific and human-geographic study designs (in the sense of workflows; contributors: 9%; members: none).

4.4 Research Objectives

The objectives for which the concept of place is being employed can be sorted into 19 categories overall. Investigating place–human relations are the most frequently articulated goals ($n = 15$), followed by the closely related categories of sense-of-place analyses ($n = 9$) and investigations into the meanings of places ($n = 8$). A more technical goal articulated often is to visualize places ($n = 8$). Overall, core members of the GIScience community showed a more technical and conceptual focus. In contrast, GIScience contributors seem to be more interested in applying place-based information and concepts to work on related applied tasks.

5 Discussion and Conclusions

The results outlined disclose interesting patterns and trends. Most GIScience authors covered in our corpus do not lay out in much detail their theoretical stance on place. Oftentimes, a range of human-geographic and philosophical authors are cited, but the actual ontological standpoint taken, especially with respect to place-based information, is not elaborated in much depth. This, on the one hand, shows that GIScience authors are aware of a range of concepts available from geography. This finding is further underpinned by our observation that core GIScience community members refer to a greater variety of place concepts than contributors coming more from the fringes of the field. On the other hand, this observation also shows that in many cases place is used in an ambiguous manner. Clearly, a more thorough understanding of place (in a geographic sense) and the nature of place-based information (in a GIScience sense) will be necessary to foster an efficient and fruitful future development of this field within GIScience.

The most frequently applied place concepts are those borrowed from humanistic geography. Accordingly, we found Yi-Fu Tuan and John Agnew to be the two most frequently cited authors, both coming from a phenomenological background rooted in experience. At the same time, humanistic geography is also concerned with the 'essence of place', i.e., the intersubjective core elements that render place important for humanity and human existence (Cresswell, 2014). This focus on the underlying structure is probably preferred by GIScience authors, because the more formal approach taken in information science is easier aligned with this viewpoint. Nevertheless, a stronger engagement with other approaches like those found in descriptive regional geography, process-oriented radical geography, or more recent relational approaches like those from non-representational geography would benefit the development of a holistic GIScience notion of place-based information.

We further found that exploratory empirical works are prevalent among the study designs applied. Being in line with our finding that there is no consensus on place in GISscience, the prevalence of exploratory approaches seems very reasonable. The given context thus favours and requires preliminary research to provide a clearer picture of what can, in principle, be done academically with the types of place-based information available. Exploratory frameworks in this sense enable the uncomplicated study of problems that lack a clear epistemological agreement caused by the ambiguity around the concept of place and what the community aims to do with it (Shields and Rangarajan, 2013). This is also well aligned with the other important type of study design, which is case studies. Looking at particular, specific places in depth, however, can still be helpful to develop a better general understanding of place.

Data-wise, GIScience research on place seems to be driven to some extent by the increasing availability of geotagged online information. In our results we found that 22% of all considered manuscripts deal primarily with one particular kind of data set in an empirical manner. These data sets includes social media but also geotagged blogs and other items from the Geoweb (Scharl and Tochtermann, 2007). Acknowledging the technical nature of the field, this can be considered a pragmatic approach. Now that new kinds of data are available, the community seems to be looking for ways to make sense of the novel kinds of information these provide. The results obtained thus indicate that the GIScience community is to some extent responding to the availability of new data sources in a pragmatic manner. Still, the GIScience community is not only driven by data but also partly by their overlap with geography. Many GIScience scholars are working in geographical institutions and thus in close contact with human geographic colleagues. This may hence explain the relatively strong interest in human-place relations including sense-of-place (in opposition to physical and other aspects of place), particularly among the core members of the community. This finding confirms a certain intellectual like-mindedness shared between GIScience and human-geographic scholars.

The results obtained are to be considered preliminary. This is due to the initial character of this review but also to the relatively small number of manuscripts ultimately considered. Still, though being non-exhaustive, the small number of manuscripts may hint on a reasonably rigorous approach. Therefore, to conclude the article, we offer the following main conclusions drawn from the evidence gained from the systematic literature review:

1. The GIScience community seems to currently focus on humanistic geographical approaches but may benefit from a more inclusive and broader consideration of geographical approaches to place.

2. The results indicate that the community is looking very much towards available data sets, thereby taking a very pragmatic viewpoint. More conceptual work independent of specific forms of data could be beneficial to the future development of the field. Some promising works into this direction are already available (e.g., Scheider and Janowicz, 2014; Winter and Freksa, 2012).

3. The review results indicate deviating behaviour in response to place between core members of GIScience and others contributing more from the fringes of the field. The work contributed by the core members thereby seems to be more conceptual in nature. This is promising as it shows that the core members of the community seem to focus on developing novel concepts of handling place-based information, while related yet more empirical scholars add another valuable viewpoint to the discourse by applying these.

Notes

1. To the best of our knowledge, the adjective *platial* has first been used by Casey (1993).

ORCID

Daniel Wagner https://orcid.org/0000-0003-2309-679X
Alexander Zipf https://orcid.org/0000-0003-4916-9838
Rene Westerholt https://orcid.org/0000-0001-8228-3814

References

Agnew, John A: *Place and politics*. Winchester, MA, USA: Allen and Unwin, 1987

—— *Space and place*. In: Agnew, John and Livingstone, David N (eds.), *The SAGE Handbook of Geographical Knowledge*, London, UK: SAGE, 2011, vol. 2011. 316–331

Blaschke, Thomas and Merschdorf, Helena: *Geographic information science as a multidisciplinary and multiparadigmatic field*. Cartography and Geographic Information Science, 41(3), 2014, 196–213. doi: 10.1080/15230406.2014.905755

Blaschke, Thomas; Merschdorf, Helena; Cabrera-Barona, Pablo; et al.: *Place versus space: from points, lines and polygons in GIS to place-based representations reflecting language and culture*. ISPRS International Journal of Geo-Information, 7(11), 2018, 452. doi: 10.3390/ijgi7110452

Borrego, Maura; Foster, Margaret J; and Froyd, Jeffrey E: *Systematic literature reviews in engineering education and other developing interdisciplinary fields*. Journal of Engineering Education, 103(1), 2014, 45–76. doi: 10.1002/jee.20038

Casey, Edward: *The fate of place: a philosophical history*, 1997

Casey, Edward S: *Getting back into place: toward a renewed understanding of the place-world*. Bloomington, IN, USA: Indiana University Press, 1993

Chen, Hao; Winter, Stephan; and Vasardani, Maria: *Georeferencing places from collective human descriptions using place graphs*. Journal of Spatial Information Science, 2018(17), 2018, 31–62. doi: 10.5311/JOSIS.2018.17.417

Cresswell, Tim: *In place / out of place: geography, ideology, and transgression*. Minneapolis, MN, USA: University of Minnesota Press, 1996

—— *Place. An introduction*. Hoboken, NJ, USA: Wiley and Sons, 2014

Curry, Michael R: *"Hereness" and the normativity of place*. In: Proctor, James D and Smith, David M (eds.), *Geography and Ethics*, London, UK: Routledge, 1999. 95–105

Davies, Clare: *Place and placing locations: a cognitive perspective*. In: Westerholt, Rene; Mocnik, Franz-Benjamin; and Zipf, Alexander (eds.), *Proceedings of the 1st Workshop on Platial Analysis (PLATIAL'18)*. 2018, 15–20. doi: 10.5281/zenodo.1472737

Drum, Peter: *Aristotle's definition of place and of matter*. Open Journal of Philosophy, 1(1), 2011, 35–36. doi: 10.4236/ojpp.2011.11006

Farinelli, Franco: *Geografia: un'introduzione ai modelli del mondo*. Turin: Einaudi, 2003

Gao, Song; Janowicz, Krzysztof; Montello, Daniel R; et al.: *A data-synthesis-driven method for detecting and extracting vague cognitive regions*. International Journal of Geographical Information Science, 31(6), 2017, 1245–1271. doi: 10.1080/13658816.2016.1273357

Goodchild, Michael F: *Geographical information science*. International Journal of Geographical Information Systems, 6(1), 1992, 31–45. doi: 10.1080/02693799208901893

—— *Twenty years of progress: Giscience in 2010*. Journal of Spatial Information Science, 2010(1), 2010, 3–20. doi: 10.5311/JOSIS.2010.1.2

—— *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*, New York, NY, USA: Springer, 2011. 21–33. doi: 10.1007/978-1-4419-7482-2_2

Lefebvre, Henri and Nicholson-Smith, Donald: *The production of space*. Chichester, UK: Wiley, 1992

Merschdorf, Helena and Blaschke, Thomas: *Revisiting the role of place in geographic information science*. ISPRS International Journal of Geo-Information, 7(9), 2018, 364. doi: 10.3390/ijgi7090364

Purves, Ross S; Winter, Stephan; and Kuhn, Werner: *Places in information science*. Journal of the Association for Information Science and Technology, 70(11), 2019. doi: 10.1002/asi.24194

Scharl, Arno and Tochtermann, Klaus: *The geospatial Web*. London, UK: Springer, 2007

Scheider, Simon and Janowicz, Krisztof: *Place reference systems. A constructive activity model of reference to places*. Applied Ontology, 9(2), 2014, 97–127. doi: 10.3233/AO-140134

Scheider, Simon and Purves, Ross: *Semantic place localization from narratives*. Proceedings of the 1st ACM SIGSPATIAL International Workshop on Computational Models of Place (COMP), 2013. doi: 10.1145/2534848.2534858

Shields, Patricia M and Rangarajan, Nandhini: *A playbook for research methods: integrating conceptual frameworks and project management*. Stillwater, OK, USA: New Forums Press, 2013

Soja, Edward W: *Thirdspace: toward a new consciousness of space and spatiality*. In: Ikas, Karin and Wagner, Gerhard (eds.), *Communicating in the Third Space*, New York, NY, USA: Routledge, 2008. 63–75

Tuan, Yi-Fu: *Space and place. The perspective of experience*. Minneapolis, MN, USA: University of Minnesota Press, 1977

Westerholt, Rene; Gröbe, Mathias; Zipf, Alexander; and Burghardt, Dirk: *Towards the statistical analysis and visualization of places*. Proceedings of the 10th International Conference on Geographic Information Science (GIScience), 2018a. doi: 10.4230/LIPIcs.GISCIENCE.2018.63

Westerholt, Rene; Mocnik, Franz-Benjamin; and Zipf, Alexander: *Introduction to the PLATIAL'18 workshop on platial analysis*. In: Westerholt, Rene; Mocnik, Franz-Benjamin; and Zipf, Alexander (eds.), *Proceedings of the 1st Workshop on Platial Analysis (PLATIAL'18)*. 2018b, 1–5. doi: 10.5281/zenodo.1475267

Winter, Stephan and Freksa, Christian: *Approaching the notion of place by contrast*. Journal of Spatial Information Science, 5, 2012, 31–50. doi: 10.5311/JOSIS.2012.5.90

Platial Geo-Temporal Demographics Using Family Names

Justin van Dijk and Paul A Longley

Department of Geography, University College London, UK

We introduce platial geo-temporal demographics as a novel way to describe places using family names as markers of migration and change at sub-national scales. By identifying the likely origins of 59,218 surnames in Great Britain, we create platial profiles of surname mixes in terms of the distance their forbears have likely migrated between 1881 and 1998/2016. By combining individual-level data derived from historic censuses of population with near-complete contemporary population registers of enfranchised adults, we demonstrate how locally and regionally distinctive surname mixes can be used in characterizing places in terms of demographic change and stasis. The results suggest that a hierarchy of places arises in Great Britain, with larger conurbations (e.g., London and Birmingham) having more surnames that can be traced back to other parts of Great Britain and beyond, as opposed to places that are characterized by the presence of a larger share of surnames that have a more local origin. These regional differences are likely linked to processes of social mobility and economic activity.

Keywords: surnames; platial profiling; geodemographics; population change

History: received on 5 July 2019; accepted on 16 August 2019; published on 27 January 2020

1 Introduction

Geodemographics has been defined as the analysis of people by where they live (Harris et al., 2005). The field relates to platial geographies in that residential neighbourhoods are in the local sense the outcome of 'birds of a feather flocking together', in ways that are replicated over entire national (and, arguably, international) settlement systems. Residential structure is thus seen as the interplay of locational proximity and social similarity played out across the urban and regional system. Geodemographics is a geography of night-time residence (Martin et al., 2015), typically measured using a melange of variables sourced from censuses and other conventional statistical or industrial sources. Some data make it possible to shift the focus from place of residence to place of work, resulting in geodemographic classifications of workplaces (e.g., Singleton and Longley, 2019). The re-use of statistics to characterize broader activity patterns than residence alone adds a temporal dimension into classification, and the term 'geo-temporal demographics' has been used to recognize that common or shared activity patterns measured over timescales ranging from the diurnal to the inter-generational can be used alongside conventional social, economic, and demographic data in order to better understand community structure. Such classifications may or may not be anchored to geographies of night-time residence. In what follows, we view inter-generational migration patterns as formative in notions of place formation.

Here we will use aggregated geographies of individuals' family names to create geo-temporal demographic platial profiles. Surnames in the British Isles came into common parlance between the 12th and 14th Centuries and have subsequently been passed down male bloodlines. Most Anglo-Saxon

J van Dijk and PA Longley (2020): *Platial Geo-Temporal Demographics Using Family Names*. In: FB Mocnik and R Westerholt (eds.), Proceedings of the 2nd International Symposium on Platial Information Science (PLATIAL'19), pp. 23–31

https://doi.org/10.5281/zenodo.3628863

PLATIAL'19 Second International Symposium on Platial Information Science (PLATIAL'19) Coventry, UK; 5–6 September 2019

family names can be described as either metonyms (pertaining to occupation, e.g., Smith and Weaver), toponyms (e.g., Hill and Gill) or diminutives (e.g., Williamson and Williams), and most have clearly identifiable geographic origins, albeit to varying levels of geographic precision (Cheshire and Longley, 2012; Longley et al., 2007). Examination of the evolving geographies of multiple surnames makes it possible to chart the outcomes of historic processes of population change and intergenerational migration (Kandt et al., 2020). From a more recent perspective, forename–surname pairings can be used to infer issues of ethnicity (Kandt and Longley, 2018) and residential segregation (Lan et al., 2018) of bearers of names that have more recently been imported from abroad. However, these foundations have not as yet been used to examine the mixes of surnames that characterize the residents of distinctive places and the cumulative effects of such mixes in the accretion and diminution of place effects.

The aim of this paper is to create data-driven Great Britain-wide platial profiles by relating the geographic origins of surnames to the locations of their bearers at different points in time. From a geodemographic perspective, this entails sifting contemporary residential structure into residents whose roots in a locality likely extend over many generations, and those whose surnames indicate bloodlines that are less established in localities. The defining characteristics of places are by no means seen as invariably grounded in inter-generational history, but our approach seeks to decompose the repetitive social similarities that characterize different locations within the national settlement system from the underlying historic structure of communities that underpins this more recent socio-spatial differentiation. Our motivation is thus to address the question: 'what place is like this place?' We conclude with some brief speculations concerning the ways in which this allows us to address issues of regional development and functional inter-dependence within the settlement system.

2 Creating Platial Surname Profiles

2.1 Data Sources

Individual-level historic census records are made publicly available 100 years after their collection date. Higgs and Schurer (2014) have brought together and standardized digital transcriptions of the censuses for England and Wales collected between 1851 and 1911 excluding 1871; data for Scotland are available for the full period 1851 to 1901. The individual census records are linked to parishes, the boundaries of which have been digitized according to two sets of consistent parish geographies. While disclosure of individual names and addresses is presently not possible from post-1911 UK censuses, a near complete set of linked addresses-level consumer registers has been assembled for the period 1997–2016 by the UK Consumer Data Research Centre (Lansley et al., 2019). This corpus of data comprises the public version of the UK Electoral Register supplemented with various consumer data sources from 2002 onwards to capture many of those that opted out of inclusion of the public version of the register or were ineligible to vote in any elections. While the linked consumer registers are incomplete and are of uncertain provenance, they have been subjected to extensive computational address matching procedures and have been partially validated by triangulation with annual Office for National Statistics' mid-year population estimates. Through these processes of internal and external validation, the Consumer Registers are found to comprise the vast majority of the UK's adult population (Lansley et al., 2019). In what follows, we examine the local, regional, national, and international origins of surnames for one historic period (1881) and two much more recent points (1998 and 2016).

2.2 Surname Origins

Using the 1881, 1998, and 2016 data we begin by locating the recorded residences of individuals bearing British surnames relative to the probable historic seed point of the name. To this end, each surname with at least 30 bearers[1] in 1851, 1861, or 1881 is assigned a geographic centroid, calculated using kernel density estimation (KDE).[2] For each surname, the isotropic fixed bandwidth is estimated using a likelihood cross-validation method, constrained by a minimum bandwidth of 5 kilometres and a maximum bandwidth of 40 kilometres. For computational reasons, we estimate the bandwidth on a sample of the surname population in cases where a surname has more than 5,000 bearers. Although, it is theoretically possible to vary the bandwidth according to the distribution of the background data to accommodate local variations, for large data sets this is extremely challenging (cf. Zhang et al., 2017). All KDE calculations are executed using the R programming language (R Core Team, 2019) and the *Sparr* package (Davies et al., 2018). To speed up processing times, calculations are parallelized using GNU Parallel (Tange, 2011) and distributed over a high-performance Linux cluster. This process resulted in 59,218 centroids, each defined as the maximum relative density value of the associated KDE for the qualifying surnames. This set is defined as our 'long-settled surname stock of British names' in the subsequent analysis (for full details on the KDE calculations and its parameters, see van Dijk et al., 2019; van Dijk and Longley, oxnd).

2.3 Spatial-Temporal Comparison

A temporally consistent zonal design is required in order to compare changing surname mixes over time. We create this by combining the consistent parishes from the historic censuses with the contemporary 2011 Office for National Statistics (ONS) Middle Layer Super Output Area (MSOA) geography. Because some of the parishes are relatively small, particularly in urban areas, we start by iteratively merging historic parishes until they reach a minimum threshold of 750 inhabitants.[3] In a second step, we assign MSOA centroids to these aggregated parishes using a point in polygon procedure, and MSOAs assigned to the same parish are merged. Parishes that do not have a single MSOA centroid within their boundaries are merged with neighbouring parishes. This procedure results in an MSOA-based temporally consistent zones (TCZs) of Great Britain consisting of 3,370 zones (cf. Kandt et al., 2020). On average each TCZ is 67 square kilometres (standard deviation: 150 square kilometres). The average population sizes of these zones in 1881, 1998, and 2016 are 8,914, 13,473 and 15,256 respectively. Once the TCZs have been defined, we create two lookup tables: one assigning MSOAs to TCZs and one assigning historic parishes to TCZs.

2.4 Origin-Destination Matrix

Next, surname migration distance is approximated using a Delaunay triangulation that connects the 3,370 TCZ centroids. Edges that are not consistent with the UK's coastal geography are removed manually. This Delaunay network is then used to create an origin–destination matrix in which each TCZ is combined with all other TCZs using Dijkstra's shortest path algorithm, in which edge length is used to measure impedance. The Delaunay triangulation is executed using the *spatstat* package in R (Baddeley et al., 2015). The creation of the origin–destination matrix is achieved through the *NetworkX* Python library (Hagberg et al., 2008).

2.5 Surname Migration Distances

Our platial profiles are based upon the mix of 'surname migration distances' of those residing within each TCZ. We thus assign each of the 59,218 long-settled surname origins to the centroid of the closest TCZ, effectively linking each surname to the origin–destination matrix. For each TCZ we then extract the individuals bearing the long-settled surnames and use the origin–destination matrix to calculate the distance that the bearers have apparently migrated from their likely origin region to their TCZ of residence. These calculations are made for 1881, 1998, and 2016. The resulting tables for 1881, 1998, and 2016 record the distances migrated by all long-term surname bearers in each TCZ.

In a last step, for every surname and TCZ combination, the log value of every migration distance is weighted by the log value of the average distance between the TCZ and all other TCZs. This yields a log-based ratio between the average distance to all other TCZs and the distance over which the bearers of the surnames have migrated. This compensates for the degree of centrality of the TCZ within Great Britain. Figure 1 shows the distribution of these values for a 1 per cent sample of the 1881 data set (about 3,000,000 rows).

2.6 Platial Profiles

The weighted distances are then classified into four origin categories: local, regional, national, and international, where the international category is comprised of surnames that do not appear in the dictionary of long-settled surnames. The classification is informed by the distribution shown in Figure 1 to estimate groups of roughly similar size. For example, surname y in TCZ k with a log-distance under

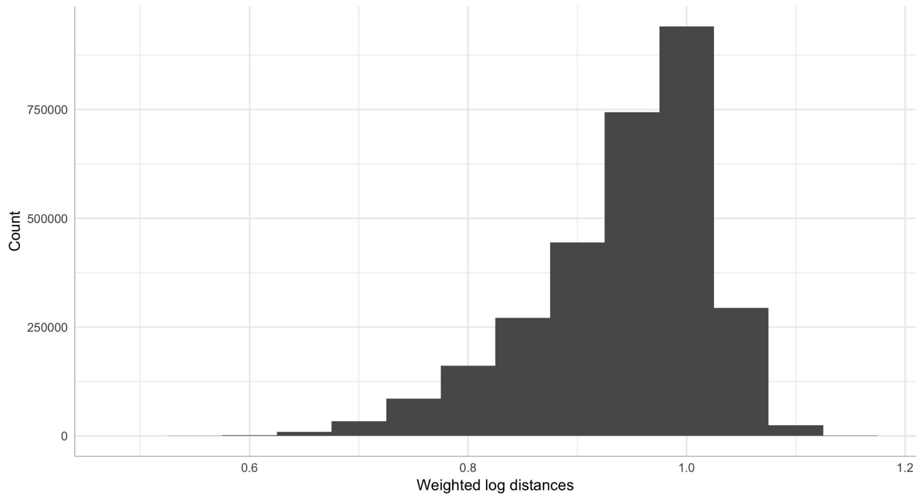


Figure 1: Weighted log-distances. Distribution of a random sample of individual long-settled surname log-distances weighted by the average log-distances per TCZ.

0.85 is considered local to this TCZ. However, this also implies that when surname y is simultaneously present in TCZ m it could have a different log-distance ratio and therefore, e.g., be considered as having a regional or wider national origin. That is, the values represent a level of locality that is relative to the TCZ centroid under consideration. Table 1 shows the distance classification used for these weighted log-distances.

3 Temporal and Regional Differences in Platial Profiles

The surname distance classification can be used to map the proportions of the population designated as having local, regional, national, or international roots. This is shown in Figure 2 for the three time periods. In 1881 areas with a high proportion of local names are found throughout the country. Particularly high proportions (over 70 per cent) of the population that are local to the area in which they reside are found in some parts of Scotland and north Wales. By contrast some major urban areas such as London and Birmingham record only 10 to 20 per cent of individuals as bearing local surnames. In 1998 and 2016, the proportions of individuals with local names have dropped significantly for all areas, indicating an increased mixture of populations.

The populations classified as having more regional roots, as presented in Figure 2b, show a similar trend over time. Other parts of Scotland and Wales have zones in which more than 70 per cent of the population is classified as regional in 1881. However, even though in general terms the regional population has decreased for most areas in 1998 and 2016, the decline is less pronounced than that of the local population. Low (below 20 per cent) proportions of individuals with regional names are found in English cities, with London being the primary exemplar.

Because the different classifications are mutually exclusive, the maps of the proportion of the population with national and international names (Figures 2c–2d) present inversed pictures of the local and regional maps. It is therefore not very surprising that the areas with the highest proportions (40 to 50 per cent) of individuals bearing national surnames in 1881 are predominantly situated in England. Rural areas typically exhibit lower proportions of surnames drawn from other parts of the country. In 1998 and 2016 it can be clearly seen that national surnames have spread throughout most

Table 1: Classification. Distance classification based on individual long-settled surname log-distances weighted by the average log-distances per TCZ.

Local	< 0.85
Regional	0.85–0.95
National	> 0.95

of the country – again manifesting the increased mixing of the long-settled population in Great Britain. London stands out as a location with a low share of individuals with national surnames (20 to 30 per cent), especially in 2016. This may seem counter-intuitive but is explicable by examining the results for international names shown in Figure 2d; by 2016 some areas in London host populations for which 60 to 70 per cent of names are not deemed to have origins in Britain.[4]

4 Discussion

Our study demonstrates that geographical surname classifications can be used to decompose local populations into local, regional, national, and international components. These classifications can be used to underpin platial geo-temporal classification of community structure. Similar to Christaller's Central Place Theory, the results cautiously suggest that a hierarchy of places exists within the British settlement system. Large conurbations, predominantly situated in England, are characterized by surnames that can be traced back to most other areas in Great Britain. London is clearly at the top of this settlement hierarchy, as identified by the high preponderance of international names. Potentially this hierarchy can be used in describing the relationship between surname diffusion and economic outcomes, such as those related to social mobility.

Some limitations of this analysis should be acknowledged. First, spatial heterogeneity in local, historic naming conventions is not fully accommodated. For example, the baseline Welsh and Scottish populations have less diverse ranges of surnames, making it possible that small increases in surnames imported from elsewhere will lead to disproportionately large apparent changes. Secondly, our surname classification is somehow arbitrary – adjusting the class boundaries of the classification will have an impact on the results. Thirdly, extremely widespread surnames (e.g., Smith) are assigned to a single origin on the basis of tiny marginal density values that bear little or no correspondence with place effects. The same applies too many other names because secondary peaks in the KDE density surface are currently not taken into consideration. Further research into this topic is likely to treat popular names, especially surnames pertaining to widely practised occupations, as having less rigidly defined origins, for instance, by employing a fluid concept of origins that defines origins at different geographical scales (cf. Kandt et al., 2020). Another solution would be to quantitatively determine the informational content of each surname, and use this informational content as a criteria for inclusion or exclusion (see Güell et al., 2018, 2015).

Notes

1. Because a KDE requires sufficient input data points, the choice for a minimum of 30 bearers is pragmatic.

2. In 1871 historic census records are only digitally available for Scotland and 1871 is therefore excluded from analysis. For surnames that have 30 or more bearers in multiple years, the earliest available year is used to determine the surname's origin.

3. Code to iteratively merge (aggregate) adjacent polygons is stored in a Postgres/PostGIS database using a minimum threshold. It is available on GitHub (https://github.com/jtvandijk/pg-polygon-merge-repo).

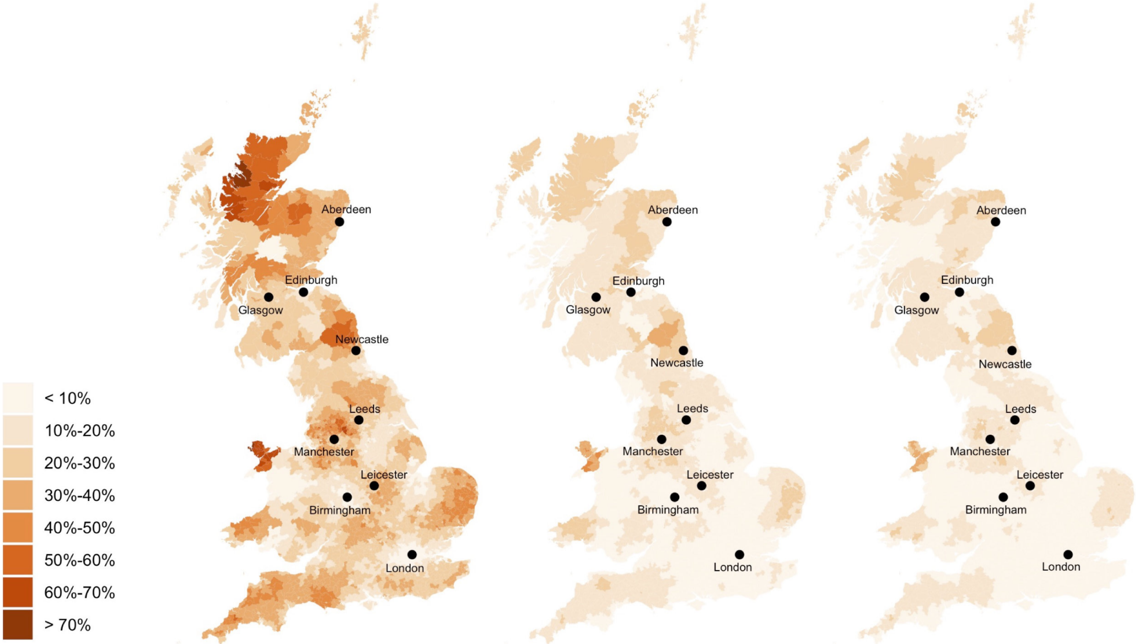
4. These surnames may nevertheless have British origins, such as surnames that had a frequency under 30 in 1851, 1861, and 1881.

Funding

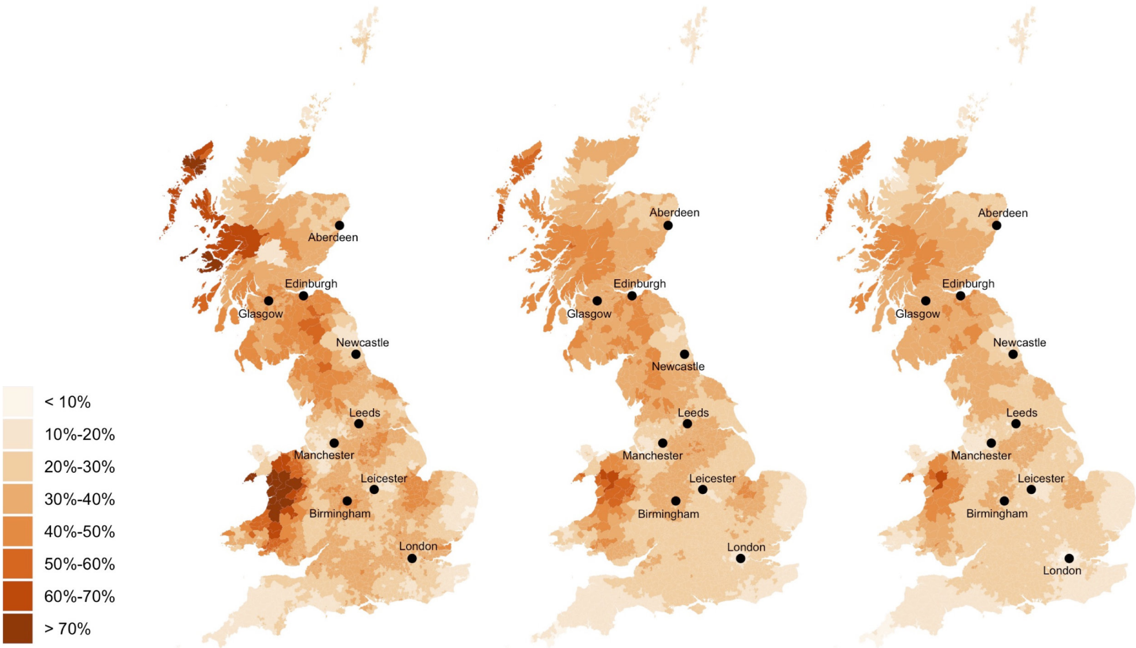
This work is supported by the UK ESRC Consumer Data Research Centre (CDRC) grant reference ES/L011840/1 and EPSRC grant EP/M023583/1 (UK Regions Digital Research Facility).

ORCID

Justin van Dijk https://orcid.org/0000-0001-5496-425X
Paul A Longley https://orcid.org/0000-0002-4727-6384

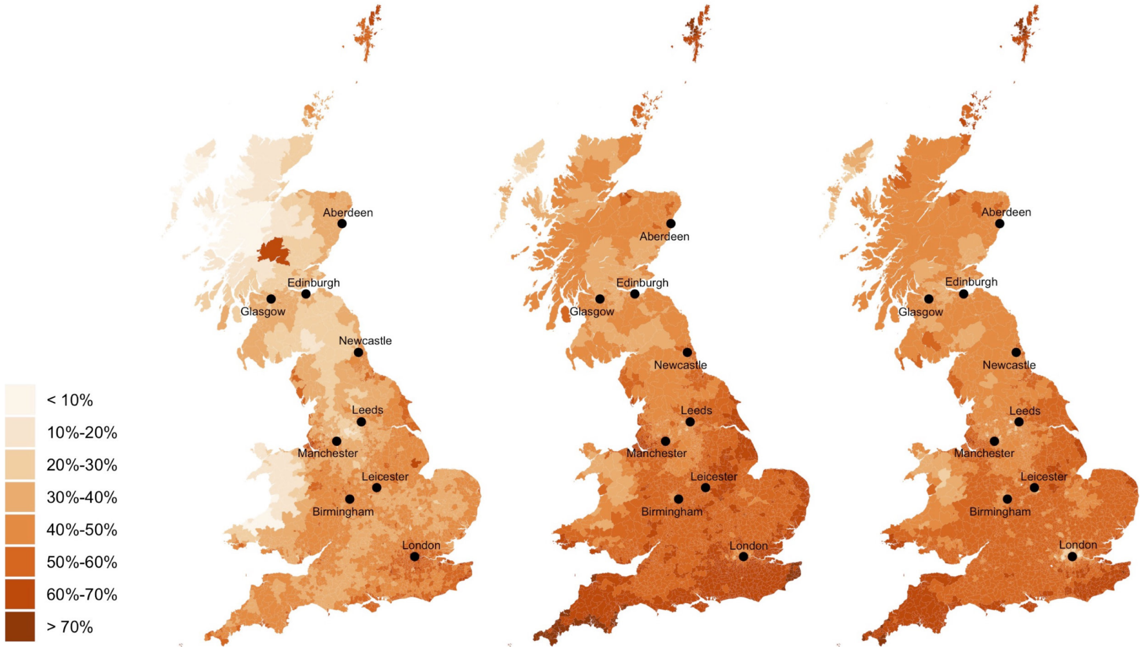


(a) British **local** roots

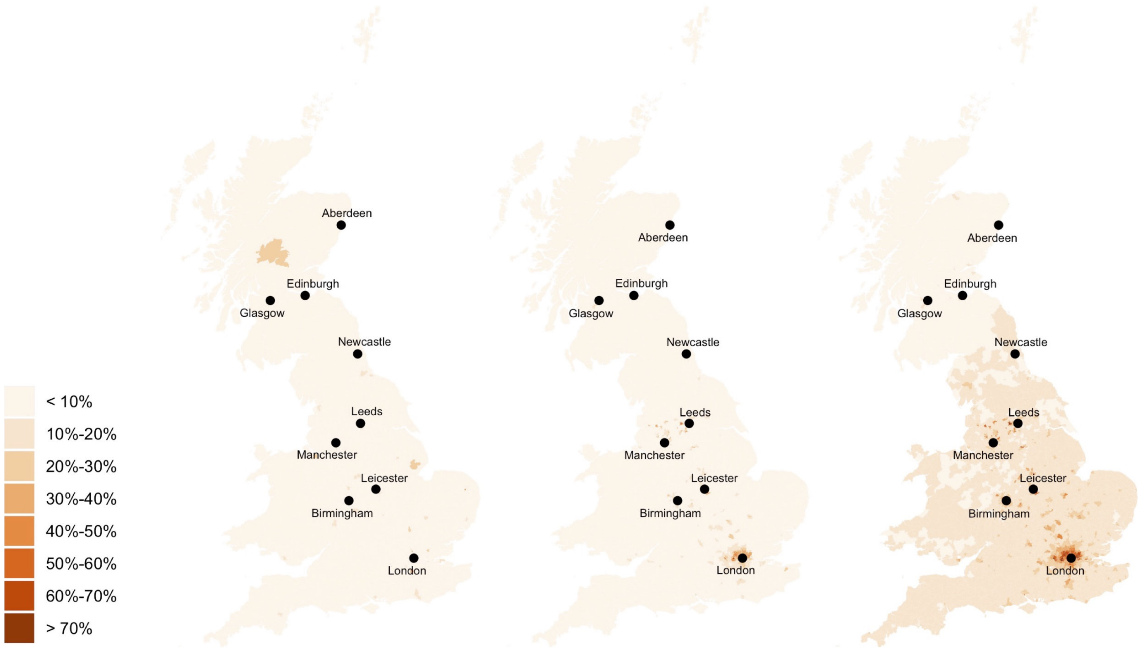


(b) British **regional** roots

Figure 2: Platial profile. Population classified by their roots in 1881, 1998, and 2016.



(c) British **national** roots



(d) British **international** roots

Figure 2 (continued): **Platial profile.** Population classified by their roots in 1881, 1998, and 2016.

References

Baddeley, Adrian; Rubak, Ege; and Turner, Rolf: *Spatial point patterns: methodology and applications with R*. London, UK: CRC Press, 2015

Cheshire, James and Longley, Paul A: *Identifying spatial concentrations of surnames*. International Journal of Geographical Information Science, 26(2), 2012, 309–325. doi: 10.1080/13658816.2011.591291

Davies, Tilman M; Marshall, Jonathan C; and Hazelton, Martin L: *Tutorial on kernel estimation of continuous spatial and spatiotemporal relative risk*. Statistics in Medicine, 37(7), 2018, 1191–1221. doi: 10.1002/sim.7577

van Dijk, Justin; Lansley, Guy; Lan, Tian; and Longley, Paul A: *Using the spatial analysis of family names to gain insight into demographic change*. Proceedings of the 27th Conference on GIS Research UK (GISRUK), 2019

van Dijk, Justin and Longley, Paul A.: *Interactive display of surnames distributions in historic and contemporary Great Britain*, n. d. Manuscript submitted for publication

Güell, Maia; Pellizzari, Michele; Pica, Giovanni; and Rodriguez Mora, Jose Vicente: *Correlating social mobility and economic outcomes*. The Economic Journal, 128(612), 2018, F353–F403. doi: 10.1111/ecoj.12599

Güell, Maia; Rodriguez Mora, Jose Vicente; and Telmer, Christopher I: *The informational content of surnames, the evolution of intergenerational mobility, and assortative mating*. The Review of Economic Studies, 82(2), 2015, 693–735. doi: 10.1093/restud/rdu041

Hagberg, Aric; Schult, Daniel; and Swart, Pieter: *Exploring network structure, dynamics, and function using NetworkX*. Proceedings of the 7th Python in Science Conference (SciPy), 2008, 11–15

Harris, Richard; Sleight, Peter; and Webber, Richard: *Geodemographics, GIS and neighbourhood targeting*. Chichester, UK: Wiley, 2005

Higgs, Edward and Schurer, Kevin: *Integrated census microdata (I-CeM), 1851-1911 [data collection]*. Data record in the UK Data Service data catalogue, 2014. doi: 10.5255/UKDA-SN-7481-1

Kandt, Jens; van Dijk, Justin; and Longley, Paul A: *Family name origins and inter-generational demographic change in Great Britain*. Annals of the American Association of Geographers, 2020, in press

Kandt, Jens and Longley, Paul A: *Ethnicity estimation using family naming practices*. PLOS ONE, 13(8), 2018, e0201774. doi: 10.1371/journal.pone.0201774

Lan, Tian; Kandt, Jens; and Longley, Paul A: *Ethnicity and residential segregation*. In: Longley, Paul A; Singleton, Alex; and Cheshire, James A (eds.), *Consumer Data Research*, London, UK: UCL Press, 2018. 71–83

Lansley, Guy; Li, Wen; and Longley, Paul A: *Creating a linked consumer register for granular demographic analysis*. Journal of the Royal Statistical Society: Series A (Statistics in Society), 182(4), 2019, 1587–1605. doi: 10.1111/rssa.12476

Longley, Paul A; Webber, Richard; and Lloyd, Daryl: *The quantitative analysis of family names: historic migration and the present day neighborhood structure of Middlesbrough, United Kingdom*. Annals of the Association of American Geographers, 97(1), 2007, 31–48. doi: 10.1111/j.1467-8306.2007.00522.x

Martin, David; Cockings, Samantha; and Leung, Samuel: *Developing a flexible framework for spatiotemporal population modeling*. Annals of the Association of American Geographers, 105(4), 2015, 754–772. doi: 10.1080/00045608.2015.1022089

R Core Team: *R: a language and environment for statistical computing*. https://www.gbif.org/tool/81287/r-a-language-and-environment-for-statistical-computing, 2019. Retrieved 27 November 2019

Singleton, Alex D and Longley, Paul A: *Data infrastructure requirements for new geodemographic classifications: the example of London's workplace zones*. Applied Geography, 109, 2019, 102038. doi: 10.1016/j.apgeog.2019.102038

Tange, Ole: *GNU Parallel – the command-line power tool*. ;login: The USENIX Magazine, 36(1), 2011, 42–47

Zhang, Guiming; Zhu, A-Xing; and Huang, Qunying: *A GPU-accelerated adaptive kernel density estimation approach for efficient point pattern analysis on spatial big data*. International Journal of Geographical Information Science, 31(10), 2017, 2068–2097. doi: 10.1080/13658816.2017.1324975

Place-Based Knowledge Systems: Human and Machine

Clare Davies

Department of Psychology, University of Winchester, UK

To tackle the problem of representing places we need to consider what people know and ask about them. This paper first suggests a list of potential queries of a 'platial' GIS, and it points out that half of them are not directly spatial but semantic, vaguely specified, and rooted in people's sensory–motor (SM) experiences of places. Recent research into human semantic cognition has indicated both explicit and separate storage of SM aspects of each stored concept or category, alongside other key features of how semantic memory appears to work. A very recent computational model taken from the semantic memory research area is briefly reviewed, which holds some potential to be applied to places. If found to effectively reflect human place cognition, this could eventually link a more humanly plausible semantic representation (probably informed by large-scale text sources) to spatial data in a GIS. Further necessary research is proposed.

Keywords: semantics; places as concepts; queries; user needs

History: received on 26 July 2019; accepted on 16 August 2019; published on 27 January 2020

1 Introduction – Querying Places

Places are not merely spaces – this much we already know. What is less clear is exactly what we need from a 'platial' system that reflects human concerns about and interests in places.

Users of spatial information are often content to get results that are location-based – what I can do, eat, see, or buy 'near here'. So what kinds of queries would be truly place-based? The author initially became involved in place research because GI users appear to need to know things that relate to ordinary people's usage of place names to denote vaguely defined and often ambiguously denoted *regions* of space, when trying to give locational information such as the address of an emergency incident. However, realistic human-understood places (as opposed to artificial administrative ones) are of concern not only to emergency services but also to data analysts exploring the attitudes, preferences, and emotions of residents and potential consumers of goods and services; to urban and rural planners attempting to improve the local environment; and to researchers trying to model the socioeconomic effects of various phenomena. What kinds of queries might such users want to make of an augmented, place-based GI system? This is an important question because, up until now, we have often talked about the need for place modelling in GI science without specifying exactly what users would need or gain from it.

Based on the kinds of specialist GIS user and their place requirements that were highlighted in previous work (Davies et al., 2009), we can hypothesize at least 30 questions which people might want to ask about places, expressed for now as natural language queries to an imaginary system. These are by no means exhaustive – further analysis of online queries such as that by Hamzei et al. (2019) may well add more, although such users were querying the general web rather than a GIS. Still, a concrete

C Davies (2020): *Place-Based Knowledge Systems: Human and Machine*. In: FB Mocnik and R Westerholt (eds.), Proceedings of the 2nd International Symposium on Platial Information Science (PLATIAL'19), pp. 33–41

https://doi.org/10.5281/zenodo.3628865

PLATIAL'19 Second International Symposium on Platial Information Science (PLATIAL'19)
Coventry, UK; 5–6 September 2019

list like the following can help us to ground discussion of what platial data needs to give us:

1. A caller said that an incident happened at the [object name or type] in [region toponym]. Which locations could they mean, and with what levels of confidence?
2. Which other locations could have been intended by [quoted name or partial address]?
3. To which place(s) is [address, location coordinates, or landmark name] most commonly said to belong?
4. How confident can we be about using [region toponym] for/at that location?
5. For whom is that toponym usage true at that point? (Any particular subpopulation? What's their frequency of visiting, and familiarity level?)
6. What's the toponym most likely to be applied at [new coordinates for a location]? (Interpolating at a point where no previous explicit data has been collected)
7. Where does [region toponym] extend to in general usage, with 90% confidence?
8. Does [any toponym] refer to more than one area, location, or object? (Disambiguation)
9. What other toponym names (local nicknames, overlapping places) are sometimes applied to [location address, coordinates, or landmark name]?
10. Show me all of the locations, objects or areas that are associated with [region toponym] with more than 80% confidence.
11. Show me only the areas and objects associated with [region toponym] by [specific subpopulation].
12. Show me a fuzzy map of (chosen area) – perhaps as understood and named by a specific subpopulation, or in a specific context of use.
13. How far do these two places [two neighbouring region toponyms] overlap?
14. How often is the same name used for both of these two [any toponyms]? By whom?
15. Where's the universally agreed core ('heart') of [region toponym]?
16. What are the main agreed characteristics of [region toponym]?
17. Which types of people feel positive, or negative, about [region toponym]?
18. How strongly do people feel about [region toponym]? (Or, show me the places where people seem to express the strongest emotions.)
19. Show me all of the places or locations where [characteristic] applies above [threshold value].
20. Show me a thematic map of [characteristic, e.g., household income] across [region toponym].
21. Show me all the places (areas and/or locations) in the city where people most often feel unsafe.
22. Show me the nearest beautiful, tranquil, lively, or busy place to [location].
23. How unsafe, busy, peaceful, or stressed do people feel in general in [region toponym], versus [another region toponym]]?
24. What types or proportions of people feel [some attitude or emotion] in [region toponym]?
25. How memorable or desirable is this [region toponym] likely to be, given its landmarks and general character (form and function)?
26. Show me a map of places that are memorable or desirable above [threshold value].
27. Which locations within [region toponym] particularly contribute to making it more memorable or desirable?

28. Tell me roughly how [region toponym]'s extents and character will appear to this type of person (someone living at specified address, someone with greater or less local familiarity).

29. Show me some places where [type of person] is likely to go.

30. Show me which places (not just specific locations within them) are generally more accessible to people with reduced mobility, or who have small children, or who have wheelchairs, buggies, or prams (strollers), or who are elderly, or who have a fear of crowds (or of open or enclosed spaces).

Note that where a toponym is included above, a place-based system needs somehow to give an answer which seems to best fit the user's context. This would ideally be based on knowledge about how toponym usage varies within a given area. In other words, disambiguation, vagueness, and context specificity are assumed to be built in and somehow 'solved' (or at least, tackled in some explicit way). Where a 'region toponym' is mentioned, this is intended to imply something like a neighbourhood or locality name at local environmental scales (Montello, 1993), but perhaps a region of a country or continent at geographic scale. Where the word 'place' is used above, it could most often be interpreted as implying a region toponym, but possibly not always. There is inherent ambiguity in what people mean when they say the word 'place', which a query system ideally would also cope with.

It will be seen that while the first fifteen queries above are spatial queries that relate mainly to places' vague, ambiguous, and overlapping extents, the second half of the list asks about various *qualities* of a region and its constituent locations. It is the latter type of question that would require more meaningful, semantic knowledge to be incorporated into a platial system.

To achieve this we may have to assume that even though each person's knowledge of a place is unique, people nevertheless will think and communicate with some degree of consensus about a place that extends across an area (i.e., subtend a region) and that many of them will recognize some of the same semantic, functional, and sensory (e.g., visible) features of a place with which they are familiar. That broad consensus knowledge will be more easily modelled in 'platial' data than the individuals' more personal and emotional associations with the place, although some level of adaptability to individual contexts is obviously also desirable in a platial information system. This paper assumes that we thus can move, carefully, from the humanistic qualitative and ontological notion of places (Cresswell, 2014; Davies, 2018) towards something that can be reconciled with GI science and systems.

Yet there is much more to this problem than merely capturing knowledge, whether commonly or individually experienced. We need to know how it is structured, and what effects this has on people's understanding and reasoning about places, so that what they 'mean' by a given toponym works in a similar way to what the system 'means' by it. Therefore, one way to bridge between humanistic and GI science understandings of places may well be the cognitive mechanisms that underpin people's platial concepts, which are at work when people reason and ask questions about them. Luckily, cognitive science has developed successively sounder, more evidence-based models of human semantic concepts and categories. Some particularly relevant recent work is outlined further below.

2 Places in Human Semantic Cognition

Places have semantic as well as spatial aspects – i.e., a place is to some extent a *concept* and, when it consists of a group of known locations within a region of any size, the place is also a *category* (of those locations). Therefore, useful learning for 'platial' modelling may emerge from current cognitive science understanding of human semantic memory and reasoning, which offers over half a century of research into our learning and processing of concepts and categories. Some initial implications of this research are explored below, and by Davies (2020).

To start with, an important aspect of learning about a new city or environment is the grouping and labelling of places. We are usually made aware quite quickly that the area has subregions within it that may be variously labelled as localities, districts, or neighbourhoods within cities, or as farms, hill ranges, conservation or wilderness areas outside them. Davies (2018, 2020) has argued that these ways of grouping locations provides cognitive efficiency, as with all categorical cognition: we can store, recall, name, reason, and communicate about multiple entities at once rather than handling each one individually.

Evidence also suggests that the way we think about region-shaped places such as urban localities follows the same tendencies and patterns as our thinking about semantic categories in general (Davies, 2020). There is *graded membership* – some members (individual locations, landmarks, or scenes) are not considered as 'obvious', 'typical', or 'good' a member as others. Furthermore, decisions about the membership of a given item (location) may vary with task context, expertise, and intent. This is the kind of finding familiar to researchers into cognition of semantic categories such as trees, birds, or household objects (Bailenson et al., 2002; Barsalou, 1985)

Theories of how we store and reason about such categories have changed and developed greatly since Rosch's famous prototype theory of the 1970s (Rosch et al., 1976). As well as hundreds of empirical experiments (for summaries see, e.g., Murphy, 2002), evidence has also been drawn from studying patients with conditions such as semantic dementia, who selectively lose their semantic knowledge while still having intact language and other skills (Patterson and Ralph, 2016; Ralph et al., 2017). This has led in recent years to an awareness that models based on simple similarity or typicality judgments and weightings amongs exemplars of a category, as might be simulated by a crude machine learning classifier, cannot explain all of the findings about how humans store semantic knowledge (Murphy, 2016).

Concepts and categories also appear to be stored in ways that take account of hierarchical membership (e.g., a robin is a bird is an animal; Raposo et al., 2012), although the hierarchy is apparently not always used when we make inferences (Murphy et al., 2012). This awareness of hierarchical structure is also found with place knowledge (Hirtle and Heidorn, 1993; Hirtle and Jonides, 1985). Human knowledge also seems to privilege a 'basic level' of a concept hierarchy: people respond more readily and use more frequently the concepts at one particular level of the hierarchy. Additionally, human categorization allows for defining crisp boundaries between categories on some occasions, for instance, where a fixed taxonomy is needed, but not in everyday casual usage. For a summary of this evidence, see Murphy (2002). There is some evidence that the basic level is also relevant in the people's understanding of the local landscape, although the study was not directly about named places (Edwardes and Purves, 2007).

Concepts are also learned from language. Indeed, for many concepts language is a primary or even the only source of information. This can also be true of places, especially ones we have not yet visited in person but have heard about verbally on repeated occasions. Thus *distributional* models of semantics, based on how names of concepts co-occur with particular sets of other words that provide explicit contexts, allow us to identify similarities between concepts whose co-occurrence patterns are similar, as judged by 'bag of words' statistical linguistic techniques such as latent semantic analysis (LSA). Such techniques tend to use huge corpora of text, rather than individual participant data, but do appear to reflect some common cognitive tendencies (Landauer, 2002).

Louwerse and Zwaan (2009) and Davies (2013) have demonstrated that where concepts are toponyms, the resulting similarity mappings closely resemble the real-world geographical layout of the places themselves, showing how even the spatial dimensions of places are implicitly represented within (large quantities of) language about them. This body of work using LSA showed that distortions in the 'map' seemed linked to people's grouping and categorizing of related places, causing them to be mentally clustered together. Some pilot human-participants work suggested that individuals' cognitive maps of their local area showed to some extent the same clustering and simplifying distortions as those derived from large-scale text corpora (Davies and Tenbrink, 2017). Thus some specific aspects of place cognition do appear to emerge from a distributional text-processing approach.

However, awareness has been increasing for 20–30 years of the 'grounding problem' in cognition: the need to link apparently abstract storage of knowledge back to our perceptual experiences of reality. 'Embodied cognition' has become a major research area over the past two decades, demonstrating that our processing even of abstract ideas often seems to involve our physical bodies (or more precisely, the areas of the brain which take inputs from our senses and which send messages to control our movements). Critiques of embodied cognition research have warned against overstating its conclusions, since 'neural reuse' (multiple purposes for each given brain area) is actually pervasive throughout the brain (Anderson, 2010). Still, it does seem true that some of our knowledge of most kinds of concept is stored in the brain areas that handle specific sensory modalities of sight, hearing, touch, smell, temperature, and taste (Ralph et al., 2017).

The 'hub and spoke' theory of semantic cognition, developed over the past 20 years, tries to account for its embodied nature (Patterson and Ralph, 2016). It argues that most concepts have an amodal

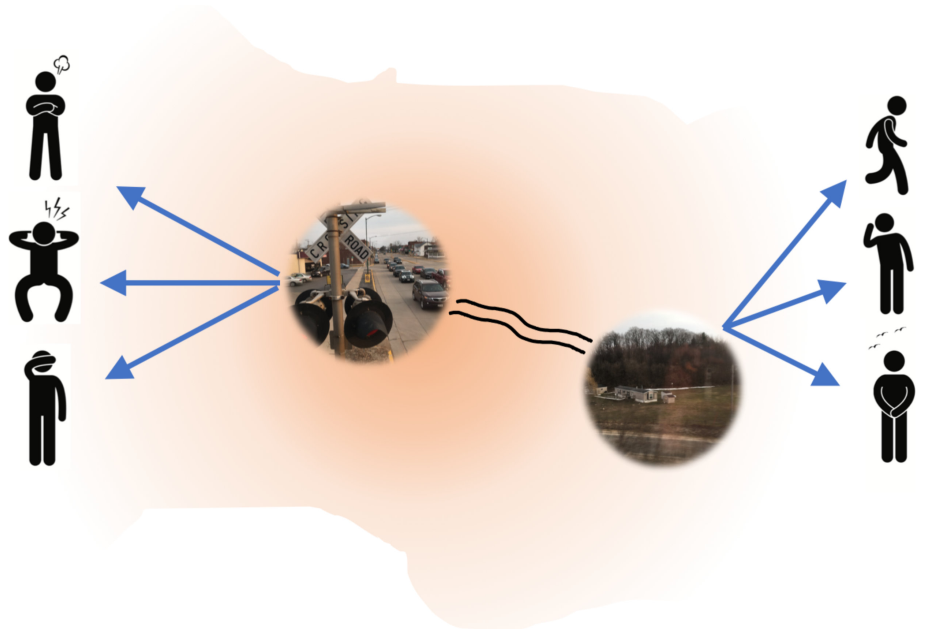


Figure 1: Applying the 'hub and spoke' theory of concepts to places. One exemplar location (scene) within a vaguely bounded locality or settlement (orange) is a railway crossing close to the centre, whose 'core' semantic knowledge is associated with certain sensory–motor properties. Being in the same town (i.e., being categorized in people's minds within that place which is named by the town's toponym) implies that to some extent, the location is linked spatially and semantically to another location on the edge of that town. But the second location may be associated with (in this case) more pleasant sensations and actions. Thus the diagram overlays the mental semantic space onto the physical geographic space, for the sake of illustration.

semantic 'core', stored in the anterior temporal cortex of the brain.[1] Importantly, that 'core' knowledge also links 'outwards' to modality-specific information about our perceptions of and motor responses to the concept, which are processed in the more specialist sensory–motor (SM) parts of the brain.

Of course, it is not of much consequence to platial information scientists that there is no single brain area where all semantic knowledge is stored. But more relevantly, the above also implies that if a place is represented mentally as a concept (and, for places beyond a single location in space, as a 'category' of contiguous locations) then it too will consist not only of a mass of similarity weightings as in the distributional models but also of separate 'embodied' knowledge of people's perceptual experiences of the place. As has already been acknowledged within the GI science literature, such knowledge should include affordances of what we can physically do at the place (Jordan et al., 1998; Kuhn, 2001) as well as sensory attributes concerning how the place looks, sounds, and feels (Mark and Freundschuh, 1995). Figure 1 attempts to apply these ideas to an imaginary person's knowledge of an example place, and of two locations (scenes) within it.

The richness of place representations will of course come as no surprise to the many qualitative researchers into 'sense of place', who have argued for the inclusion of such qualities for many years (e.g., Tuan, 1977). But it finally draws a clear three-way link between such research, more controlled empirical evidence, and potentially useful computational models of semantic cognition, the most relevant of which will be outlined briefly below. Thus these may finally start to be of direct use for 'place' modelling.

3 Modelling Meaningful Places Computationally

The semantic, but also sensory–motor experiential, nature of places implies that we will not fully capture people's knowledge of them simply by attributing locations within a GIS to one locality versus another – even if done using graded 'fuzzy' membership, so that locations at the margins have their ambiguous affiliation recognized (e.g., Brindley et al., 2018). Embodied cognition also implies that even adding known semantic attributes to such records of places within a GIS, perhaps via ontologies (Fonseca et al., 2002), even if those ontologies formally attempt to capture SM 'affordances' in terms of functions and activities that can be performed at the place (Kuhn, 2001), *still* may not adequately reflect the way that people consider them, or the structure of our knowledge about them. Can we use large-corpus data to extract place knowledge that reflects the structure of human semantic memory, while also adding a sensory–motor 'grounding' to our place data?

One possible way forward has been proposed, based on accumulated cognitive science evidence (Hoffman et al., 2018). Hoffman et al. constructed an interesting prototype connectionist computational model of semantic knowledge, based on the above 'hub and spoke' insight that we explicitly store sensory–motor attributes of concepts, plus separately abstracted 'core' knowledge. The model, like the long-established PDP (parallel distributed processing) model that preceded it, uses layers of hidden units to achieve representation of concepts according to their attributes (e.g., bird 'has wings' and 'can fly'). Unlike previous models, though, this one also learns from both distributional data (co-occurrence patterns of words with other words), and the temporal (sequential) context in which the words appear (since the proximity of a word or name to particular other words is also an important indicator of its semantics). The model also explicitly encodes the sensory–motor properties of some of the more concrete words in the input stream, so as to gradually learn to associate those properties with the names of related concepts. Thus in terms of modelling place, the inclusion of a park that was widely referenced as beautiful, or was heavily used for physical exercise, might cause those SM-related attributes to become associated with the overall neighbourhood. Since people speak and write of sensory–motor attributes of a place as well as more objective, measurable ones, it seems quite plausible to capture at least some of them from text sources.

Some of the goals of Hoffman et al.'s model were concerned specifically with the learning of concepts from language, and the inputs made to the system in its experimental version appear to have been artificially constructed rather than learning from naturally occurring discourse. As yet, it is also unclear whether the model replicates the various other facets of human semantic cognition mentioned earlier – the hierarchy with its preferred 'basic level', the fuzzy boundaries that are occasionally crisp, etc. – although its predecessor models did so and thus this may be tentatively assumed. Yet the inclusion of SM properties does start to give this kind of model the potential to model concepts that are cognitively realistic, with an explicit focus on what people physically experience and do.

It is also worth noting, of course, that locations and places are often also learned sequentially – not only from verbal information but also from travelling through and around them along linear paths through the space. Thus the sequentially sensitive aspect of the model is of interest for place learning from direct experience, as well as from verbal information.

Overall, this kind of hybrid approach seems to have the potential for such a model to effectively model places *as people understand them*, and to be linked to a GIS or other information system to allow for queries to be made about not only where places extend to and what data and language is associated with them, but also what they are like – beginning to model the 'sense' of place at last. Even so, of course, many questions remain to be answered about whether and how efficiently such models can be expanded for a real-world application rather than a small research demonstration, and also about how they can be linked to existing GIS. It is also as yet unclear whether such a model can replicate *all* of the aspects of category-based reasoning that may be relevant to places, since to date not all of those aspects have even been established empirically – an ongoing research project in itself. As suggested by one reviewer of this paper, it is also as yet unclear whether a system based upon the architecture that appears to underpin *individual* cognition can actually improve a model of (place) information that is inevitably collated across *populations*, in its aim to address multiple user needs.

4 Conclusion and Future Research

This paper has tried to outline some specific ways in which queries to a 'platial' GIS might go beyond the spatial and formally semantic. It then reviewed briefly some of the key findings in human semantic memory, and a recent model that seems to hold promise for modelling people's place knowledge computationally, such that a system could eventually result that truly reflected people's place understanding.

Obviously, a next stage should involve attempting to reproduce such a system, in which places in a specific local area would be the concepts it learned about from some large-scale text sources (probably web scraping). One would then test its outputs against human-participant experiments, and thence (if it appears to work) examine the important question of scaling up the system for more extensive geographies. These are not small challenges, but they may finally bring us closer to having genuinely 'platial' GIS.

Notes

1. Interestingly, this is quite close to the parahippocampal areas known to be involved in spatial processing of scenes and navigating large-scale spaces (Epstein, 2008), which may in fact be more generally involved in the processing of the semantic context associated with any stimulus according to Aminoff et al. (2013).

Acknowledgements

Thanks to FB Mocnik for his initial suggestion to compile a list of potential 'platial' system queries.

ORCID

Clare Davies https://orcid.org/0000-0003-0261-2353

References

Aminoff, Elissa M; Kveraga, Kestutis; and Bar, Moshe: *The role of the parahippocampal cortex in cognition*. Trends in Cognitive Sciences, 17(8), 2013, 379–390. doi: 10.1016/j.tics.2013.06.009

Anderson, Michael L: *Neural reuse: a fundamental organizational principle of the brain*. Behavioral and Brain Sciences, 33(4), 2010, 245–266. doi: 10.1017/S0140525X10000853

Bailenson, Jeremy N; Shum, Michael S; Atran, Scott; Medin, Douglas L; and Coley, John D: *A bird's eye view: biological categorization and reasoning within and across cultures*. Cognition, 84(1), 2002, 1–53. doi: 10.1016/S0010-0277(02)00011-2

Barsalou, Lawrence W: *Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories*. Journal of Experimental Psychology: Learning, Memory, and Cognition, 11(4), 1985, 629–654. doi: 10.1037/0278-7393.11.1-4.629

Brindley, Paul; Goulding, James; and Wilson, Max L: *Generating vague neighbourhoods through data mining of passive web data*. International Journal of Geographical Information Science, 32(3), 2018, 498–523. doi: 10.1080/13658816.2017.1400549

Cresswell, Tim: *Place. An introduction*. Hoboken, NJ, USA: Wiley and Sons, 2014

Davies, Clare: *Reading geography between the lines: extracting local place knowledge from text*. Proceedings of the 11th International Conference on Spatial Information Theory, 2013, 320–337. doi: 10.1007/978-3-319-01790-7_18

—— *Place and placing locations: a cognitive perspective*. In: Westerholt, Rene; Mocnik, Franz-Benjamin; and Zipf, Alexander (eds.), *Proceedings of the 1st Workshop on Platial Analysis (PLATIAL'18)*. 2018, 15–20. doi: 10.5281/zenodo.1472737

—— *Places as fuzzy locational categories*. Acta Psychologica, 2020, 102937. doi: 10.1016/j.actpsy.2019.102937

Davies, Clare; Holt, Ian; Green, Jenny; Harding, Jenny; and Diamond, Lucy: *User needs and implications for modelling vague named places*. Spatial Cognition and Computation, 9(3), 2009, 174–194. doi: 10.1080/13875860903121830

Davies, Clare and Tenbrink, Thora: *Place as location categories: learning from language*. Proceedings of the 13th International Conference on Spatial Information Theory, 2017, 217–225. doi: 10.1007/978-3-319-63946-8_37

Edwardes, Alistair J and Purves, Ross S: *A theoretical grounding for semantic descriptions of place*. Proceedings of the 7th International Symposium on Web and Wireless Geographical Information Systems, 2007, 106–120. doi: 10.1007/978-3-540-76925-5_8

Epstein, Russell A: *Parahippocampal and retrosplenial contributions to human spatial navigation*. Trends in Cognitive Sciences, 12(10), 2008, 388–396. doi: 10.1016/j.tics.2008.07.004

Fonseca, Frederico T; Egenhofer, Max J; Agouris, Peggy; and Camara, Gilberto: *Using ontologies for integrated geographic information systems*. Transactions in GIS, 6(3), 2002, 231–257. doi: 10.1111/1467-9671.00109

Hamzei, Ehsan; Li, Haonan; Vasardani, Maria; et al.: *Place questions and human-generated answers: a data analysis approach*. Proceedings of the 22nd AGILE Conference on Geographic Information Science, 2019, 3–19. doi: 10.1007/978-3-030-14745-7_1

Hirtle, Stephen C and Heidorn, P Bryan: *The structure of cognitive maps: representations and processes*. In: Gärling, Tommy and Golledge, Reginald G (eds.), *Advances in Psychology: Behavior and Environment*, 1993. 170–192. doi: 10.1016/S0166-4115(08)60043-6

Hirtle, Stephen C and Jonides, John: *Evidence of hierarchies in cognitive maps*. Memory and cognition, 13(3), 1985, 208–217. doi: 10.3758/BF03197683

Hoffman, Paul; McClelland, James L; and Ralph, Matthew A Lambon: *Concepts, control, and context: a connectionist account of normal and disordered semantic cognition*. Psychological Review, 125(3), 2018, 293–328. doi: 10.1037/rev0000094

Jordan, Troy; Raubal, Martin; Gartrell, Bryce; and Egenhofer, Max J: *An affordance-based model of place in GIS*. Proceedings of the 8th International Symposium on Spatial Data Handling, 1998, 98–109

Kuhn, Werner: *Ontologies in support of activities in geographical space*. International Journal of Geographical Information Science, 15(7), 2001, 613–631. doi: 10.1080/13658810110061180

Landauer, Thomas K: *On the computational basis of learning and cognition: arguments from LSA*. In: Ross, Brian H (ed.), *Psychology of Learning and Motivation*, London, UK: Elsevier, 2002, vol. 41. 43–84. doi: 10.1016/S0079-7421(02)80004-4

Louwerse, Max M and Zwaan, Rolf A: *Language encodes geographical information*. Cognitive Science, 33(1), 2009, 51–73. doi: 10.1111/j.1551-6709.2008.01003.x

Mark, David M and Freundschuh, Scott M: *Spatial concepts and cognitive models for geographic information use*. In: Nyerges, Timothy L; Mark, David M; Laurini, Robert; and Egenhofer, Max J (eds.), *Cognitive Aspects of Human-Computer Interaction for Geographic Information Systems*, Dordrecht: Springer, 1995. 21–28. doi: 10.1007/978-94-011-0103-5_3

Montello, Daniel R: *Scale and multiple psychologies of space*. Proceedings of the 1st International Conference on Spatial Information Theory (COSIT), 1993, 312–321. doi: 10.1007/3-540-57207-4_21

Murphy, Gregory L: *The big book of concepts*. Cambridge, MA, USA: MIT Press, 2002

—— *Is there an exemplar theory of concepts?* Psychonomic Bulletin and Review, 23, 2016, 1035–1042. doi: 10.3758/s13423-015-0834-3

Murphy, Gregory L; Hampton, James A; and Milanovic, Goran S: *Semantic memory redux: an experimental test of hierarchical category representation*. Journal of Memory and Language, 67(4), 2012, 521–539. doi: 10.1016/j.jml.2012.07.005

Patterson, Karalyn and Ralph, Matthew A Lambon: *The hub-and-spoke hypothesis of semantic memory*. In: Hickock, Gregory and Small, Steven L (eds.), *Neurobiology of Language*, London, UK: Elsevier, 2016. 765–775. doi: 10.1016/B978-0-12-407794-2.00061-4

Ralph, Matthew A Lambon; Jefferies, Elizabeth; Patterson, Karalyn; and Rogers, Timothy T: *The neural and computational bases of semantic cognition*. Nature Reviews Neuroscience, 18, 2017, 42–55. doi: 10.1038/nrn.2016.150

Raposo, Ana; Mendes, Mafalda; and Marques, J Federico: *The hierarchical organization of semantic memory: executive function in the processing of superordinate concepts*. NeuroImage, 59(2), 2012, 1870–1878. doi: 10.1016/j.neuroimage.2011.08.072

Rosch, E; Simpson, C; and Miller, RS: *Structural bases of typicality effects*. Journal of Experimental Psychology: Human Perception and Performance, 2(4), 1976, 491–502. doi: 10.1037/0096-1523.2.4.491

Tuan, Yi-Fu: *Space and place. The perspective of experience*. Minneapolis, MN, USA: University of Minnesota Press, 1977

Shared Mental Models as a Psychological Explanation for Converging Mental Representations of Place – the Example of OpenStreetMap

Maren Mayer[1], Daniel W Heck[1], and Franz-Benjamin Mocnik[2,3]

[1]*School of Social Sciences, University of Mannheim, Germany*
[2]*Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, the Netherlands*
[3]*Institute of Geography, Heidelberg University, Germany*

People perceive the environment in various idiosyncratic ways, letting them conceptualize places differently. Representation in a data set and communication about places, however, create the need to reach agreement in the place a symbol or word represents. People have thus to integrate their views about a place. In this paper, we discuss how idiosyncratic views about places and their integration can be traced in OpenStreetMap. Then, we explore novel ways of how to model the integration processes of such idiosyncratic views by the means of psychological models. In particular, we explore the concept of Shared Mental Models. Such formal modelling and the corresponding better understanding of how people integrate their views about places improves the way we can make sense of collaborative shared data sets.

Keywords: concepts of place; shared mental models; collaborative behaviour; knowledge integration; ontology merging; OpenStreetMap (OSM)

History: received on 16 July 2019; accepted on 16 August 2019; published on 27 January 2020

1 Introduction

Where in Coventry does the International Symposium on Platial Information Science take place? This question will most likely be answered differently by the various participants and conveners ahead of the symposium because they have different ways to get to the symposium and because they have different experiences about Great Britain. If this question is asked at the end of the symposium, however, the participants would provide more similar answers. Many of them possibly would, e.g., include the Zeeman Building in their answer, i.e., the building in which the symposium takes place. This expected difference in how participants describe the place where the symposium takes place can be explained in terms of the mental model that each of the participants has. These mental models differ because the participants have different backgrounds and experiences. They conceptualize differently. During the symposium, the participants stay in the same room, they interact, and they have possibly similar ideas about why they participate. By this shared experience their mental models can become entangled and thus closer aligned, which is why they describe the place in a potentially more similar way than before.

Social interaction can reveal how individuals align their mental models about places. Places are at least in parts socially constructed (e.g., Kyle and Chick, 2007; Low and Altman, 1992), as is also the case in the introductory example. When contributing to OpenStreetMap (OSM), a major example of Volunteered Geographic Information (VGI; Goodchild, 2007) and of Geographical Shared Data

M Mayer, DW Heck, FB Mocnik (2020): *Shared Mental Models as a Psychological Explanation for Converging Mental Representations of Place – the Example of OpenStreetMap*. In: FB Mocnik and R Westerholt (eds.), Proceedings of the 2nd International Symposium on Platial Information Science (PLATIAL'19), pp. 43–50

https://doi.org/10.5281/zenodo.3628871

PLATIAL'19 Second International Symposium on Platial Information Science (PLATIAL'19)
Coventry, UK; 5–6 September 2019

Sources (Mocnik et al., 2019), individuals compare their own mental model of a place to the representation they find in the data or on the map, potentially leading to an adaption of the representation. Thereby, it needs to be kept in mind that the way individuals read the map is strongly affected by the way information is visually represented (Keil et al., 2018). Also, the information conveyed by a map is strongly limited with respect to places (Mocnik and Fairbairn, 2018), which is a particular reason for why map reading is strongly influenced by our mental models. The folksonomy of OSM is a result of a shared representation. Individuals semantically annotate elements in a certain way, which is then adopted by others in a social process. This way of how the elements are annotated is documented in the OSM Wiki, but this documentation is merely a representation of the social process and its results (Mocnik et al., 2017). The differences in the way individuals cognitively experience and conceptualize places and the geographical environment can be traced in the data, as has been done in the case of social fragmentation and corresponding cartographic epistemologies in Jerusalem (Bittner, 2014, 2017; Bittner and Glasze, 2018).

The paper is structured as follows. First, we discuss examples of how individuals represent places in idiosyncratic ways in OSM data (Section 2). Tracing these conceptualizations in the data is not straightforward – places are represented differently because they are different. Such differences in how we conceptualize can, however, be revealed when several representations of the same place at different points in time are compared (Section 3). As these differences do not coexist in the current version of the data (but only in their history), they need to be integrated. Such integration can be conceptualized by referring to the mental model that the contributors have (Section 4). Shared mental models, an approach from psychological research, can be used to understand and model how contributors decide for or against adapting the representation of places in OSM data (Section 5). Future research may use the proposed theoretical and methodological advances for explaining and predicting future inconsistencies in how we represent places in socially created data sets.

2 Examples of Idiosyncratic Representations of Places

Places are cognitively constructed by perceiving and making sense of what exists in space (Malpas, 1999; Tuan, 1977). The way we perceive and cognitively process our environment is subject to systematic distortion (Tversky et al., 1999). We thus conceptualize the same environment in different ways (Freundschuh and Kitchin, 1999; Tversky et al., 1999), leading to different ideas of the places we are in. For instance, we tend to align compass directions with the axes of the maps we use (Tversky, 1981). Sketch maps provide means to examine in detail the various ways in which individuals conceptualize places (Boschmann and Cubbon, 2014). Classical maps are, unlike sketch maps, different in the sense of that they usually adhere to space and geometries rather than to our perception. These maps provide the illusion of objectivity, which is despite the fact that the use of different map projections leads to very different geometrical representations. Further, maps offer the advantage that they can be created largely independently of the viewer and according to intersubjective scales, at least when compared to sketch maps.

Data sets and projects that form part of VGI are the result of a social creation process, in which usually a large number of individuals are involved. Many OSM elements have, e.g., been edited by several contributors, who potentially conceptualize places differently. These various conceptualizations lead to potentially differing representations. Some of the differences might no longer be visible in the data or the map representation because the folksonomy used is not fine-grained enough. Differences in water temperature, e.g., are so far not represented – there exists no corresponding OSM key. Other differences are, however, visible in the data. As an example, single trees are represented as such in the OSM data set while forests are often represented as 'tree-covered areas'. Contributors might judge in different ways whether a group of trees shall be mapped as single trees or as a tree-covered area. Accordingly, cognition and the way we conceptualize places has an impact on mapping behaviour.

Individuals experience a separation between public and private space, in particular in Arabic cultures (Costa and Noble, 1986). This difference, often referred to as public/private space dichotomy, can be traced in OSM. In Figure 1, a part of the city Ras Al Khaimah in the United Arab Emirates is depicted. The blocks of houses are interrupted by streets. Within a block of houses, no systematic gaps are visible on the satellite imagery (Figure 1a). The corresponding part of OSM, however, shows an elongated gap. The contributors have added a retail area that is received as public space and another

(a) Satellite imagery **(b)** OpenStreetMap

Figure 1: Representation of public and private space in Ras Al Khaimah, United Arab Emirates. In the centre part of (b), the map shows a gap between public retail space and private space (indicated by red arrows). This gap is most likely the result of the experienced separation between public and private space, because the satellite imagery (a) does not reveal any spatial structure alike. Copyright of map data: (a) Google Earth, DigitalGlobe, (b) OpenStreetMap contributors (cf. www.openstreetmap.org/copyright)

area that is received as private space (Figure 1b). The public/private space dichotomy is a seemingly plausible explanation for the gap between these two areas, but the validity of this explanation cannot be proven in retrospect.

3 Tracing Conceptualizations in the OpenStreetMap History

The various conceptualizations of place coexist in the OSM data set. A geographical feature can, e.g., be represented in one way while another feature is represented in a different way. Also, the representation of a geographical feature can change over time. That is, a contributor can change the geometric or thematic representation of a geographical feature in order to adapt it to his or her own conceptualization. As an entity can, however, only be represented in one way at a time, the contributors have to finally agree on a representation. That is, they have to find a coherent conceptualization of a place in the context of the used folksonomy, and they have to align their implicit assumptions about space. This situation is different to conventional data sources in the context of Geographic Information Systems (GIS), the latter of which are not able to capture places (Couclelis, 2005; Freundschuh and Egenhofer, 1997; Goodchild and Li, 2011). It seems, however, to be important to understand the ways individuals conceptualize places and negotiate these, because the conceptualization of a place underlying a map has an influence on the cartographic representation.

The history of a data set is the key to tracing conceptualizations. When examining the representation of a geographical feature, the conceptualization by the contributor who added the feature can at least to some degree be traced. Such knowledge would, however, not reveal how the conceptualization relates to the geographical feature, i.e., how the contributor distorts certain geometric and thematic features, unless ground truth is known. In order to minimize the effort and ensure that the results are of high quality, an intrinsic assessment would be favourable.[1] The different ways contributors conceptualize a feature can be traced intrinsically by comparing how one contributor created a representation of a geographical feature and how another one changed this representation. The history of the data, which is available in case of OSM,[2] is needed to examine such changes and accordingly the various representations of the feature.

4 Conceptualizing the Integration of Idiosyncratic Representations

Every person holds an idiosyncratic mental representation of their surrounding environment including geographical and geometrical features as well as information about properties, affordances, and functions (Freundschuh and Kitchin, 1999; Tversky et al., 1999). Nevertheless, the OSM community integrates these individual mental representations into a single shared conceptualization represented in the data. Many different processes such as the motivation to contribute, decision making on what to add or to change, and the emergence of shared editing rules among contributors can affect the development of the data. The resulting OSM data contain a shared representation formed among the contributors. Thus, each entry, either as history or currently used information, can be considered as a behavioural outcome of the processes that formed this shared representation.

In a first step, methods adapted from multiattribute decision making (Heck et al., 2017) could be applied to OSM data in order to understand in which situations users tend to adapt existing representations. Research on multiattribute decision making examines how individuals consider the attributes of available options, e.g., technical properties of different smart phones, and how these attributes contribute to the decision made. This notion could be adapted to contributions to the OSM data set by examining which attributes or settings are relevant for the decision whether to adapt the currently used representation (e.g., in case of non-rectangular corners, coarse mapping of buildings and streets, and missing tags). The obtained findings about this decision process can later be verified through a laboratory experiment. In such a setting, OSM data with different levels of mapping and tagging accuracy can be presented to participants who then decide whether they want to adapt the fictitious representation or not. This experiment can help to gain deeper insights into the preconditions of changes within OSM data.

Even though multiattribute decision making theories focus on individuals, analysing the OSM data with these methods offer the opportunity to apply these individual centred theories to a collaboration setting. This may extend theories on multiattribute decision making, may help to further understand decision making in social contexts as well as convergence among contributors in a decision context.

5 Shared Mental Models – a Potential Explanation for Converging Data

An understanding of how individuals make sense of the environment by forming a mental representation of entities and their relationships has been approached in the past (Johnson-Laird, 1980). The concept of mental models formalizes this idea by defining mental models as cognitive structures that resemble the structure of what is represented and allow logical connections between entities within the mental models (Johnson-Laird, 1980, 2005). When information changes or new information is available, mental models are adjusted accordingly (Johnson-Laird, 2005).

Research about mental models mostly concentrated on individual processes, in particular reasoning, by examining the logical inferences participants draw from given statements (Johnson-Laird, 2005). However, work and organizational psychology adapted this idea to group processes of team members working together in organisations, called shared mental models (Converse et al., 1991; Klimoski and Mohammed, 1994). Among several theories on how shared representations among team members foster team performance (Akkerman et al., 2007; Klimoski and Mohammed, 1994; Turner et al., 2014), the idea of shared mental models (Converse et al., 1991), sometimes also referred to as team mental models (Klimoski and Mohammed, 1994; Mohammed et al., 2010), could serve as a theoretical basis for research on OSM data as well. Shared mental models are mental models that overlap to a certain degree between team members (Converse et al., 1991). This overlap is conceptualized by the *degree of detail* and *accuracy*, the latter of which describes the extent to which the shared mental model captures the real-world environment. Other authors only differentiate between similarity as the degree of the sharedness of the model among team members and accuracy (Mohammed et al., 2010). In the context of organizations, shared mental models cover all facets of task completion, such as which tasks are performed, how these are performed and by whom, and which resources are available for completing a task.

To examine whether shared mental models are applicable, it needs to be tested whether contributors actually hold shared mental models. These models are expected to comprise contribution rules about mapping and tagging. They can be measured in surveys by asking contributors to describe how they would add or change certain objects to the data set. Since shared mental models are not only characterized by their content but also by the structure of their content, specific methods can be applied to measure these structural components along with the content (DeChurch and Mesmer-Magnus, 2010). To measure the accuracy of the measured shared mental models, the OSM Wiki can help to create a baseline model as a standard for comparison.

After exploring to what extent mental models are shared by the contributors, further research could examine the integration process of individual mental models leading to the development of shared mental models in more detail. While research on shared mental models often addresses the consequences of shared mental models on the performance of a team inside an organization (Mohammed et al., 2010; Turner et al., 2014), OSM data offer the opportunity to examine how shared mental models emerge – every element can be traced back to its creation in the data set. Research already demonstrated that shared mental models are the result of an extensive communication process (Van den Bossche et al., 2011). Especially, constructive conflicts, i.e., the discussion of differences in understanding between team members, is positively associated with the extent of similarity of mental models in teams. This communication process seems to be facilitated through the possibility to use visualization during the communication (Bittner and Leimeister, 2014). Unfortunately, however, OSM contributors mostly communicate only indirectly, which limits the examination of shared mental models. Nevertheless, hypotheses derived from the concept of shared mental models about how the contributed information is adapted over time, how contribution behaviour changes, and which contributors remain active within the community can still be tested. Constructive conflict, although not verbally, can emerge when contributors correct one another. As this correction is less interactive and less comprehensive as verbal communication and as it lacks any direct visualization, it might not suffice to change the mental models of other contributors. While some contributors may change their contributing behaviour and thus further participate in the development of a shared mental model, others may refrain from contributing when being corrected frequently.

6 Conclusions

Combining Geographical Information Science and Psychology to explore contribution behaviour to the OSM data set can lead to a deeper understanding of the relationship between idiosyncratic mental representations and shared representations of places in the data. Analysing the relative importance of feature properties for the decision whether and how to change the feature in question can lead to further insights on how geographical features are represented by contributors. Furthermore, it could be expected that, due to continuous changes of the representation, the contribution behaviour of all contributors converges over time. The OSM data include the representation of geographical features and their history, thereby potentially reflecting the assumed mental models. Extracting the mental model from the data seems, however, to be challenging because one cannot properly distinguish between the mental model, conventions of how to represent, and properties of the represented feature. The analysis of the data should therefore be complemented by experiments under controlled conditions and in a laboratory setting. These experiments would, e.g., examine whether the mental models of the contributors shift over time, or whether only the contribution behaviour and thus the representation within the OSM data changes. Findings obtained from the theoretical and methodological advances proposed in the present paper may help to explain current and predict future inconsistencies in representing places in the long run.

Notes

1. For a discussion of the ways to intrinsically assess the data, see the corresponding publication by Mocnik et al. (2018a).

2. For a computational examination of the history of OSM data, the framework *OSHDB* (Raifer et al., 2019) can be used. Further OSM-related data sources can provide additional context (Mocnik et al., 2018b).

Author Contributions

M Mayer, DW Heck, and FB Mocnik contributed the main idea jointly. M Mayer wrote Sections 4–6, while FB Mocnik wrote Sections 1–3. The writing and conceptualization of the paper was coordinated and supervised by FB Mocnik.

Funding

M Mayer has been supported by the Heidelberg Academy of Sciences and Humanities, and by the DFG graduate school *Statistical Modeling in Psychology* (GRK 2277). DW Heck has been supported by the DFG graduate school *Statistical Modeling in Psychology* (GRK 2277). FB Mocnik has been supported by the DFG project *A framework for measuring the fitness for purpose of OpenStreetMap data based on intrinsic quality indicators* (FA 1189/3-1).

ORCID

Maren Mayer https://orcid.org/0000-0002-6830-7768
Daniel W Heck https://orcid.org/0000-0002-6302-9252
Franz-Benjamin Mocnik https://orcid.org/0000-0002-1759-6336

References

Akkerman, Sanne; Van den Bossche, Piet; Admiraal, Wilfried; et al.: *Reconsidering group cognition: from conceptual confusion to a boundary area between cognitive and socio-cultural perspectives?* Educational Research Review, 2(1), 2007, 39–63. doi: 10.1016/j.edurev.2007.02.001

Bittner, Christian: *Reproduction of social fragmentations in OpenStreetMap: the example of Jerusalem*. Kartographische Nachrichten, 64(3), 2014, 136–144

—— *Diversity in volunteered geographic information: comparing OpenStreetMap and Wikimapia in Jerusalem*. GeoJournal, 82(5), 2017, 887–906. doi: 10.1007/s10708-016-9721-3

Bittner, Christian and Glasze, Georg: *Excluding effects of cartographic epistemologies – thinking about mapping paradigms in OpenStreetMap and Wikimapia*. Kartographische Nachrichten, 3, 2018, 120–125

Bittner, Eva Alice Christiane and Leimeister, Jan Marco: *Creating shared understanding in heterogeneous work groups: why it matters and how to achieve it*. Journal of Management Information Systems, 31(1), 2014, 111–144. doi: 10.2753/MIS0742-1222310106

Boschmann, E Eric and Cubbon, Emily: *Sketch maps and qualitative GIS: using cartographies of individual spatial narratives in geographic research*. The Professional Geographer, 66(2), 2014, 236–248. doi: 10.1080/00330124.2013.781490

Van den Bossche, Piet; Gijselaers, Wim; Segers, Mien; Woltjer, Geert; and Kirschner, Paul: *Team learning: building shared mental models*. Instructional Science, 39(3), 2011, 283–301. doi: 10.1007/s11251-010-9128-3

Converse, Sharolyn A; Cannon-Bowers, Janis A; and Salas, Eduardo: *Team member shared mental models: a theory and some methodological issues*. Proceedings of the Human Factors Society Annual Meeting, 35(19), 1991, 1417–1421. doi: 10.1177/154193129103501917

Costa, Frank J and Noble, Allen G: *Planning Arabic towns*. Geographical Review, 76(2), 1986, 160–172. doi: 10.2307/214622

Couclelis, Helen: *Space, time, geography*. In: Longley, Paul A; Goodchild, Michael F; Maguire, David J; and Rhind, David W (eds.), *Geographical information systems: principles, techniques, management and applications*, New York: Wiley, 2005. 2nd edn., 29–38

DeChurch, Leslie A and Mesmer-Magnus, Jessica R: *Measuring shared team mental models: a meta-analysis.* Group Dynamics: Theory, Research, and Practice, 14(1), 2010, 1–14. doi: 10.1037/a0017455

Freundschuh, Scott M and Egenhofer, Max J: *Human conceptions of spaces: implications for GIS*. Transactions in GIS, 2(4), 1997, 361–375. doi: 10.1111/j.1467-9671.1997.tb00063.x

Freundschuh, Scott M and Kitchin, Rob: *Contemporary thought and practice in cognitive mapping research: an introduction*. The Professional Geographer, 51(4), 1999, 507–510. doi: 10.1111/0033-0124.00187

Goodchild, Michael F: *Citizens as sensors: the world of volunteered geography*. GeoJournal, 69(4), 2007, 211–221. doi: 10.1007/s10708-007-9111-y

Goodchild, Michael F and Li, Linna: *Formalizing space and place*. Proceedings of the International Colloquium on Founding Territorial Sciences (CIST), 2011, 177–183

Heck, Daniel W; Hilbig, Benjamin E; and Moshagen, Morten: *From information processing to decisions: formalizing and comparing psychologically plausible choice models*. Cognitive Psychology, 96(1), 2017, 26–40. doi: 10.1016/j.cogpsych.2017.05.003

Johnson-Laird, Philip N: *Mental models in cognitive science*. Cognitive Science, 4(1), 1980, 71–115. doi: 10.1207/s15516709cog0401_4

—— *Mental models and thought*. In: Holyoak, Keith J and Morrison, Robert G (eds.), *The Cambridge handbook of thinking and reasoning. Thinking and reasoning: a reader's guide*, Cambridge, UK: Wiley, 2005. 185–208

Keil, Julian; Mocnik, Franz-Benjamin; Edler, Dennis; Dickmann, Frank; and Kuchinke, Lars: *Reduction of map information regulates visual attention without affecting route recognition performance*. ISPRS International Journal of Geo-Information, 7(12), 2018, 469. doi: 10.3390/ijgi7120469

Klimoski, Richard and Mohammed, Susan: *Team mental model: construct or metaphor?* Journal of Management, 20(2), 1994, 403–437. doi: 10.1016/0149-2063(94)90021-3

Kyle, Gerard and Chick, Garry: *The social construction of a sense of place*. Leisure Sciences, 29(3), 2007, 209–225. doi: 10.1080/01490400701257922

Low, Setha M and Altman, Irwin: *Place attachment. A conceptual inquiry*. In: Altman, Irwin and Low, Setha M (eds.), *Place attachment*, New York, NY: Plenum, 1992. 1–12. doi: 10.1007/978-1-4684-8753-4_1

Malpas, Jeff E: *Place and experience. A philosophical topography*. Cambridge, UK: Cambridge University Press, 1999

Mocnik, Franz-Benjamin and Fairbairn, David: *Maps telling stories?* The Cartographic Journal, 55(1), 2018, 36–57. doi: 10.1080/00087041.2017.1304498

Mocnik, Franz-Benjamin; Ludwig, Christina; Grinberger, A Yair; et al.: *Shared data sources in the geographical domain—a classification schema and corresponding visualization techniques*. ISPRS International Journal of Geo-Information, 8(5), 2019, 242. doi: 10.3390/ijgi8050242

Mocnik, Franz-Benjamin; Mobasheri, Amin; Griesbaum, Luisa; et al.: *A grounding-based ontology of data quality measures*. Journal of Spatial Information Science, 16, 2018a, 1–25. doi: 10.5311/JOSIS.2018.16.360

Mocnik, Franz-Benjamin; Mobasheri, Amin; and Zipf, Alexander: *Open source data mining infrastructure for exploring and analysing OpenStreetMap*. Open Geospatial Data, Software and Standards, 3(7), 2018b. doi: 10.1186/s40965-018-0047-6

Mocnik, Franz-Benjamin; Zipf, Alexander; and Raifer, Martin: *The OpenStreetMap folksonomy and its evolution*. Geo-spatial Information Science, 20(3), 2017, 219–230. doi: 10.1080/10095020.2017.1368193

Mohammed, Susan; Ferzandi, Lori; and Hamilton, Katherine: *Metaphor no more: a 15-year review of the team mental model construct*. Journal of Management, 36(4), 2010, 876–910. doi: 10.1177/0149206309356804

Raifer, Martin; Troilo, Rafael; Kowatsch, Fabian; et al.: *OSHDB: a framework for spatio-temporal analysis of OpenStreetMap history data*. Open Geospatial Data, Software and Standards, 4(3), 2019. doi: 10.1186/s40965-019-0061-3

Tuan, Yi-Fu: *Space and place. The perspective of experience*. Minneapolis, MN: University of Minnesota Press, 1977

Turner, John R; Chen, Qi; and Danks, Shelby: *Team shared cognitive constructs: a meta-analysis exploring the effects of shared cognitive constructs on team performance*. Performance Improvement Quarterly, 27(1), 2014, 83–117. doi: 10.1002/piq.21163

Tversky, Barbara: *Distortions in memory for maps*. Cognitive Psychology, 13(3), 1981, 407–433. doi: 10.1016/0010-0285(81)90016-5

Tversky, Barbara; Morrison, Julie Bauer; Franklin, Nancy; and Bryant, David J: *Three spaces of spatial cognition*. The Professional Geographer, 51(4), 1999, 516–524. doi: 10.1111/0033-0124.00189

Affective Route Planning Based on Information Extracted from Location-Based Social Media

Madalina Gugulica, Eva Hauthal, and Dirk Burghardt

Institute of Cartography, TU Dresden, Germany

Research on environmental psychology has shown that humans perceive environments not solely by relating to their physical features but also in reference to their affective qualities. Without regard to the impact these affective perceptions might have on the decision-making process during navigation tasks, most of the current web or mobile pedestrian routing services employ distance or time optimized algorithms and often fail to fully meet the needs of the users. This study wishes to introduce a pragmatic approach of harnessing Location-Based Social Media as a potent and readily available resource for extracting people's affective perceptions of the environment. Furthermore, it proposes a method to aggregate and model the extracted information for the enhancement of pedestrian route planning and identifies the main issues and challenges the nature of these data raises.

Keywords: affective perception; location-based social media content; pedestrian route planning

History: received on 18 July 2019; accepted on 16 August 2019; published on 27 January 2020

1 Introduction

Emotion influences most aspects of cognition and behaviour in the human-space interaction. Hence, humans perceive space and environments not only according to their physical features but also affectively (Bondi et al., 2006; Russell, 2003; Ulrich, 1983). Emotions are a critical aspect of how people experience places (Mody et al., 2009). Thus, capturing and studying the affective meanings people attach to places may lead to a better comprehension of people's spatial experiences and behaviour, as well as to the enhancement of certain location-based services (e.g., navigation systems, travel applications, or urban planning) by offering them a social-context.

In the last decade, numerous researchers have been interested in improving pedestrian navigation systems. Significant progress has been achieved concerning the design of human-centred route planning algorithms to compute more intuitive routes meeting other criteria than just distance and duration. For instance, in order to meet the users' need of simplicity in navigation, Winter (2002) proposed the computation of the 'best' routes in terms of routes with minimal angles and number of turns, while Haque et al. (2007) proposed the calculation of routes minimizing the number of complex intersections with turn ambiguities. Also, the level of safety a user might experience has been the focus of several studies about pedestrian navigation. Miura et al. (2011) consider the illumination of streets to be an important aspect with respect to the level of safety experienced. Similarly, Bao et al. (2017) proposed a pedestrian routing algorithm that considers the street illumination and the width of the sidewalks as significant safety indicators. Recently, Novack et al. (2018) proposed the computation of 'pleasant routes' by quantifying the attractiveness of a route in terms of presence of green areas, social places, and noise. Their customized pedestrian routes are entirely based on OpenStreetMap[1] data.

M Gugulica, E Hauthal, and D Burghardt (2020): *Affective Route Planning Based on Information Extracted from Location-Based Social Media*. In: FB Mocnik and R Westerholt (eds.), Proceedings of the 2nd International Symposium on Platial Information Science (PLATIAL'19), pp. 51–60

https://doi.org/10.5281/zenodo.3628877

PLATIAL'19 Second International Symposium on Platial Information Science (PLATIAL'19)
Coventry, UK; 5–6 September 2019

Most of the approaches aiming to improve route planning for pedestrians are, however, based on objective, physical parameters. To the best of our knowledge, the *AffectRoute* project (Huang et al., 2014) was the first work to acknowledge the importance of considering the subjective relations between people and places to navigation applications. The authors suggest that in order to meet the users' needs concerning safety or attractiveness, people's affective responses to the environment should be integrated in the route computation, which is why they proposed a crowd-sourcing approach to collect the affective information. The results of their empirical evaluation of the algorithm showed that routes ('AffectRoutes') generated considering people's affective perceptions of environments were preferred over the conventional shortest ones. Nevertheless, even with the growing popularity of crowd-sourcing in the era of Web 2.0, collecting enough data for a successful implementation of the routing algorithm at a larger scale was fairly challenging, since this acquisition method requires intense user involvement. Consequently, it is important to find additional meanings of how to collect these subjective data and this is what the present study aims for.

With the increasing use of social media networks (e.g., Twitter[2], Instagram[3], and Flickr[4]) and the dissemination of mobile devices equipped with positioning sensors, Location-Based Social Media (LBSM) data have become a valuable stream of information in environmental and geographical analysis. As a result of the availability, economy and the potential to include more diverse, more detailed, more local, and more contextualized data, mining LBSM data has gained significant attention in the last decade. Topics such as landmarks and movement patterns discovery from geotagged images (Jankowski et al., 2010), place-semantics extraction (Hollenstein and Purves, 2010), context-aware location recommendations (Huang, 2016), studying the people's perception and knowledge of environments (Huang et al., 2013), or extraction of space related emotions (Hauthal and Burghardt, 2016) exemplify some of the research directions this particular type of data has ushered in.

Following the initiated research pathways, this study regards LBSM data, namely the metadata of geotagged Flickr and Instagram images, as a potent and readily available resource for extracting people's affective perceptions of places and environments, and it proposes a methodology to harness these data for pedestrian route planning purposes.

It is important to mention that the methodology treats the affective perceptions in LBSM as taking the shape of linguistic expressions, which are coded with the affective connotations of the words the user chooses to describe the content shared. Moreover, the term *affective perception* encompasses two other concepts, namely *affective response* and *affective quality*. Affective response defines the general internal neuropsychological state of an individual that occurs as an emotional reaction to a single or to various stimuli, while affective quality is the emotion-inducing quality that people verbally attribute to a place, often expressed by describing words such as 'beautiful', 'ugly', 'peaceful', 'hectic', 'boring', etc. (Mehrabian and Russell, 1974). To quantify the affective perceptions, we adopted Russell's dimensional approach (Russell and Pratt, 1980) of classifying and quantifying affective states, which characterizes emotions along two primary dimensions of arousal and valence. Arousal is the extent to which a stimulus is calming or exciting, whereas valence represents the extent to which a stimulus is unpleasant (negative) or pleasant (positive). According to these two perceived dimensions, emotions can be classified by plotting the valence and arousal on the two-dimensional affective space.

2 Approach to Extract and Aggregate the Affective Perceptions

To test the methodology developed, we chose to work with Flickr and Instagram content, since the georeferenced photographs, in most cases, depict places and the information contained in the metadata most likely refer to the captured places. We used metadata (titles, descriptions, tags, and geographical coordinates) attached to Flickr and Instagram images uploaded between 2007 and 2018. The data set comprises 137,421 Flickr and 708,123 Instagram geotagged images located within the boundaries of the city of Dresden.

The approach proposed is based on two assumptions. First, the content that people share on Flickr and Instagram often contains emotional expressions about the transited places, expressions that are not explicit but, rather, coded with affective connotations of the words people choose to describe the content shared. Thus, for instance, a photograph of a place the user perceives as attractive and comfortable would be described and tagged using words with a positive affective connotation (e.g., 'beautiful', 'safe', and 'clean'), whereas the words to describe a less attractive and uncomfortable environment would have a rather negative affective connotation (e.g., 'dirty', 'noisy', and 'desolated'). Secondly, we assume that people often include a place name (e.g., 'Zwinger', 'Frauenkirche', and 'Albertstraße') or a place noun (e.g., 'museum', 'church', or 'street') in the title, description, or set of tags of their content when referring to the perceived environment. The presented approach, in a rather heuristic manner, considers only the metadata that complies with this criterion.[5] To decode the affective perceptions, lexicon-based sentiment analysis was performed on the collected metadata, while a proximity analysis was used to aggregate (in this case, by averaging) the captured affective perceptions in relation to a navigation graph. A detailed description of the conceptual work is presented in the following.

2.1 Data Preprocessing

Location-based social media data come with different levels of positional accuracy depending on the geotagging method that was used. For Flickr, geotags may be retrieved as geographic coordinates from the EXIF (Exchangeable Image File Format) metadata of the image uploaded (if existent), provided by the users either by means of synchronization with track logs from the built-in GPS of their devices or by manually locating photos using a map interface. Flickr stores in the metadata of each geotagged image a location accuracy level ranging from 1 (world level) to 16 (street level), which is automatically assigned depending on the precision of the GPS coordinates or the zoom level of the map used to locate the image. Instagram also offers the users the possibility to geotag their content; however, they are constrained to the selection of the location from a pre-defined list of platial tags hosted by Facebook, which is based on the EXIF location metadata (if available) or the current location of their device. Until August 2015, users were also able to add custom place-labels based on the EXIF metadata coordinates or the device's position. Depending on the geotagging procedure used, a geotagged image does not always provide the exact geographic coordinates of the location from which the image was taken, a fact that leads to a series of issues concerning the positional accuracy and quality of the data. This represents one of the main disadvantages of working with LBSM data and needs to be addressed in future work.

Since the purpose of this work is to integrate the extracted affective responses with a routing graph, a street positional accuracy level is required. Thus, the entries of the data that did not comply with these requirements need to be filtered out. For Flickr data, only geotagged images with a level of positional accuracy higher than 14 or classified under the name 'street' were selected to be further analyzed. Regarding the Instagram data, this process was more complicated due to the lack of information concerning the positional resolution. When searching for a practical solution, we used the platial tags provided by the Instagram API and manually filtered out all the entries with platial tags representing large areas such as the entire city (e.g., 'Dresden') or districts of the city (e.g., 'Loschwitz' and 'Strehlen'). Nevertheless, this solution needs to be revised and improved, because it is based on a rather superficial evaluation of the positional accuracy of the geotagged photos and might lead to inaccuracies in the analysis. In addition, to avoid user-biased valence and arousal values, the metadata generated by a single user at a certain location was merged and duplicates were removed.

Text Normalization, Language Detection, and Hashtags Segmentation. One of the main disadvantages of using LBSM data is that the textual content has a low degree of formal semantic and syntactic accuracy. In order to provide only significant information for the lexicon-based sentiment analysis to be performed, all the hyperlinks, mentions, emoticons, numbers, and other characters besides letters and punctuation were removed. Furthermore, automatic language identification was applied to each entry via the *polyglot*[6] Python library, a natural language pipeline that supports multilingual applications. Very short sentences (< 5 characters) and language predictions of other languages than English and German or with a low confidence (< 0.5) were excluded from the results. According to the language identified, the misspelling and typing errors were corrected using the *pyspellchecker*[7] Python library.

Despite the abundance of noise inherent in social media data, trending hashtags often contain important information and provide insight into the users' affective evaluations of places and environments, and the affective states experienced. Hashtag segmentation is, therefore, an important first step in natural language processing (NLP), since it would be difficult to derive accurate meaning from a piece of text without initially determining the words it consists of. For the extraction of single words from multi-word hashtags, an algorithm was employed that recursively breaks the substrings of the input and searches for the longest matching word in a list of English[8] or German[9] words (depending on the language detected in the former step) until it returns a meaningful phrase.

2.2 Extraction of the Affective Perceptions

To extract the affective perceptions from the metadata of the geotagged images, the lexicon-based sentiment analysis or keywords spotting method was used. Lexicon-based sentiment analysis employs NLP techniques to tokenize, tag the parts-of-speech, and lemmatize text into a list of words, which are further looked up in affective lexicons to determine their valence and arousal values. The approach computes the affective connotation of adjectives and nouns. The reason behind this choice is that while adjectives usually represent the affective qualities of places in adjective-noun constructions (e.g., 'beautiful building', 'quite place', and 'noisy area'), single nouns could also be regarded as an expression of the affective qualities perceived (e.g., 'this place is a ruin', 'the beauty of this area'). Adjectives in verb-adjective constructions (e.g., 'I feel comfortable', 'I feel happy', 'I am scared') represent expressions of affective states as well. The affective connotations of the single words were quantified with values for the dimensions of valence and arousal according to the English affective lexicon created by Warriner et al. (2013) and the BAWL-R (The Berlin Affective Word List Reloaded) German affective lexicon (Võ et al., 2009). The polarity classification of the BAWL-R lexicon was employed, which rates valence on a 7-point scale ranging from -3 (very negative) through 0 (neutral) to +3 (very positive) and arousal on a 5-point scale ranging from 1 (low activation) to 5 (high activation).

In the algorithm developed to preprocess each metadata entry, every single lemmatized word was looked up in the corresponding affective lexicon and, in case it matched one of the lexicon's entries, it was stored in a database together with the corresponding valence and arousal values. Due to the multilingual nature of social media data and the fact that language identification was initially performed at sentence level (for performance reasons – detecting the language of single words has lower levels of accuracy), language identification was performed anew on single words. If the word was identified as belonging to the other language considered and a match in the corresponding lexicon was found, its corresponding affective dimensions as well as the word itself were added to the database. Otherwise, a list of synonyms and hypernyms was generated, and each of them was looked up in the affective lexicon until it matched one of the entries. The word was skipped in case no match was detected. After iterating through all tokens of an entry, the overall affective polarity of the metadata and the average value of the arousal were computed as mean values.

2.3 Aggregation and Modelling of the Affective Perceptions

The basic idea behind this phase is to aggregate the extracted individual affective perceptions, model them as collective affective perceptions, and encode them as a weighting attribute for each street segment of a navigation graph. In this way, the encoded collective affective ratings can be integrated in a typical cost function to compute and deliver routes that consider, apart from physical features such as time and distance, also the affective perceptions towards places and environments.

Spatial Extent of the Affective Perceptions and Aggregation Method. As an aggregation method we propose the computation of the arithmetic mean value of the valence and arousal dimensions of all captured affective perceptions, which are relevantly located with respect to the street segment. In our approach we follow the idea propounded by Pippig et al. (2013) and define the perceptual spatial extent as the idealized space, the features of which could be affectively perceived when located on a specific street segment. Out of practical reasons and in order to enable the implementation of the approach, the proposed methodology regards the perceptual spatial extent as the space represented by the close proximity and limits it to a 100 m buffer area around each street segment. Nevertheless, it is problematic to define a general extent to which a place can be perceived affectively. Further research needs to be conducted in this sense. Since our choice is not based on empirical evidence, we applied the same reasoning on which the Inverse Distance Interpolation method is based. Accordingly, we assume each affective perception extracted to have a local influence that diminishes with distance. Consequently, in the computation of the arithmetic mean value of the valence and arousal values for each street segment, the discrete values closest to the street segments receive greater weights. The weights diminish as a function of distance.

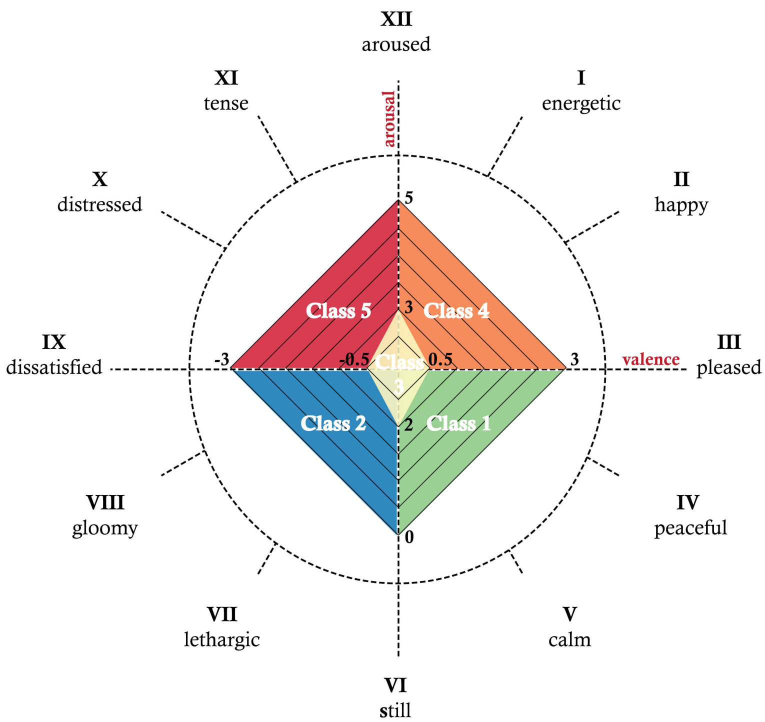


Figure 1: Five affective classes. Adapted from the 12-point circumplex structure of the core affect (Yik et al., 2011, p. 706)

Quantification and Modelling of the Affective Perceptions. For the quantification and classification of the affective ratings, we applied a bipolar framework that is organized around the dimensions of valence and arousal and is largely based on the circumplex model of affect (Russell and Pratt, 1980). The framework employs a single bipolar scale ranging from positive to negative to measure the subjective valence. The changes in reported valence are subjectively different from the changes in the reported arousal. Therefore, this framework also employs an additional arousal scale ranging from low (sleepy/inactive) to high (excited/active). As can be observed in Figure 1, the different emotional states can be represented at any level of valence and arousal, or at a neutral level of one or both of these factors. In order to classify the extracted affective perceptions by following the same principle, five possible affective classes were conceptualized as follows:

1. *Positive valence and low arousal affective class (valence* > 0.5 *and arousal* < 2.5*).* In this case the affective responses could reflect a pleasant, comfortable, calm, or attractive environment such as gardens or parks.

2. *Negative valence and low arousal affective class (valence* < -0.5 *and arousal* < 2.5*).* This class represents places that evoke sentiments of sadness and dissatisfaction. For example, war memorials or places associated with sad events.

3. *Neutral valence and neutral arousal affective class (*$-0.5 \leq$ *valence* ≤ 0.5 *and* $2 \leq$ *arousal* ≤ 3*).* Due to the numerous affective ratings that have slightly positive or negative connotations (values between 0 and ± 0.5), this approach proposes the mentioned interval $[-0.5, 0.5]$ as neutral valence values. Moreover, it was observed that most of the affective responses that fall in the category of neutral valence have arousal values between 2 and 3. This class includes rather ordinary places that do not particularly activate pedestrians affectively. An example of such an environment could be represented as common residential areas.

4. *Positive valence and high arousal affective class (valence* > 0.5 *and arousal* > 2.5*).* This would be the case of energetic and pleasant places, such as busy touristic areas of a city, sport centres, and shopping areas.

5. *Negative valence and high arousal affective class (valence < −0.5 and arousal > 2.5).* This class is representative for areas and places that evoke feelings of fear, terror or, generally, a high level of discomfort. Some examples could be represented by infamous suburbs where the rate of criminality is high, sites of unpleasant events such as attacks and revolts, or noisy areas with intensive traffic.

3 Results

To test the methodology described, an algorithm was developed and tested on the trial data. Consequently, 243,882 individual affective perceptions, which were generated by 98,560 distinct users, could be captured. The visualization of the affective perceptions (see Figure 2) revealed one of the main issues LBSM data raise, namely the irregular spatial distribution of the users' contributions caused by the popularity bias. Thus, data peaks emerge in central areas and highly frequented tourist hot-spots, while peripheral or less frequented areas are marginalized.

Before classifying the individual affective perceptions extracted and in order to ensure the validity of the classes conceived, five lists of prevailing words for each of them were created. Inspecting the word lists, the presence of a relatively high number of representative affective terms confirmed that the valence and arousal values ranges chosen were suitable. The classification results showed that most of the extracted affective ratings (71% of the total) belong to the *positive valence and low arousal affective class*. The second most prevalent affective class (with 19%) was the one represented by neutral valence and arousal values, while the other three classes summed up a total of just 8% of the whole

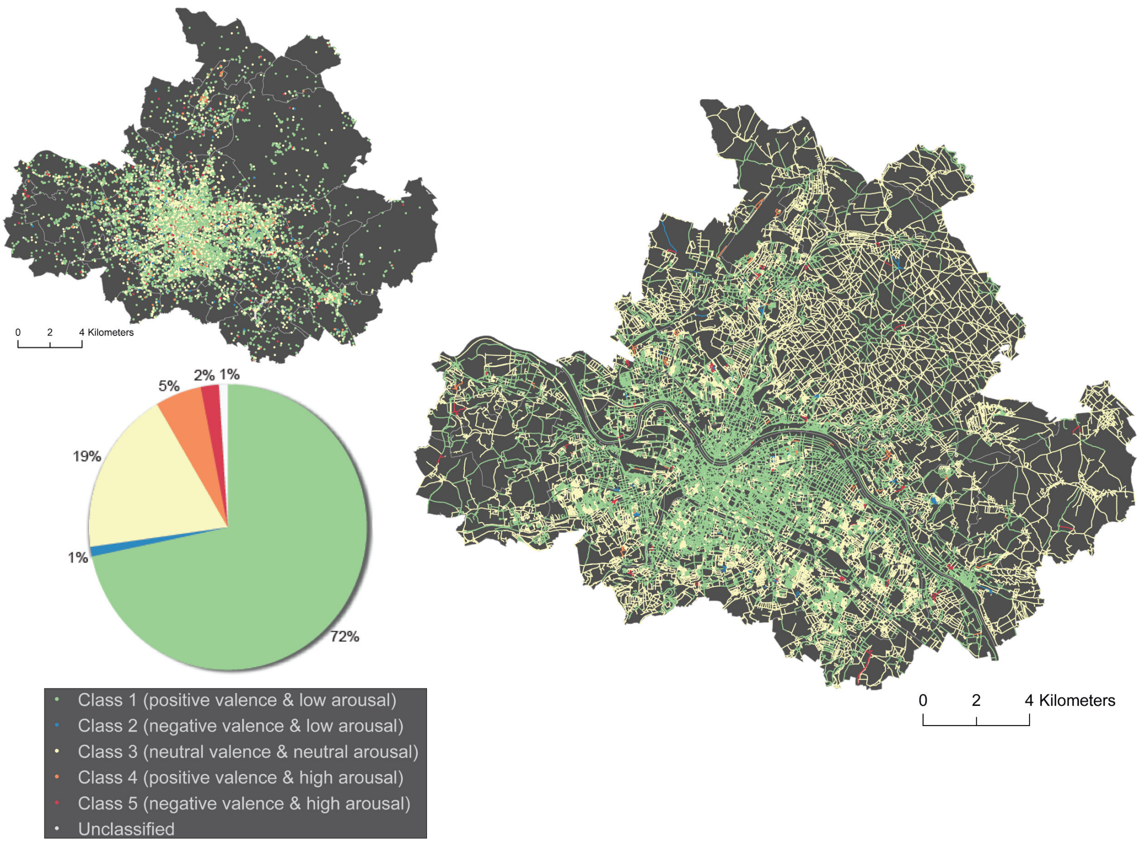


Figure 2: Classification of the street segments according to the extracted affective perceptions. The visualization of the classified extracted affective perceptions (left map) reveals an irregular spatial distribution as well as an uneven distribution of the classes. The impact of these results on the classification of the street segments (right map) is twofold: 40% of the street segments got assigned a local average valence and arousal due to the sparsity of the identified affective perceptions and over 50% of the street segments are characterized by positive valence and low arousal.

extracted affective perceptions and 2% remained unclassified – the ranges selected for the classification not covering their affective dimensions. This might indicate that social media users have a general tendency to share content with a positive or neutral affective association, such as pictures of popular, pleasant places that evoke positive emotions, or images that depict common environments encountered on a daily basis, such as the areas where the users live, work, or study.

The classification of the street segments, according to the aggregated collective ratings, follows the same trend. More than half of the edges of the graph have been assigned positive valence and low arousal weights. At a small difference, the next class that comprises a large part of the edges is represented by the *Neutral valence and arousal affective class* (nearly 49% of the network). However, the data sparsity issue also has repercussions on the classification of the street segments. While some of the streets have been associated numerous affective ratings, some others have no associated ratings at all and get assigned a local average valence and arousal of all the affective perceptions identified in the corresponding district. Thus, only for 10% of the edges, the aggregated valence and arousal values rely on the discrete affective perceptions. For almost 40% of the street segments, there were no data available. The other three affective classes are representative for less than 1% of network, whereas 0.36% of the edges remain unclassified because the valence and arousal values do not fall within any of the proposed intervals.

4 Discussion and Conclusion

Throughout the conducted research a considerable number of issues and difficulties concerning the nature of LBSM data could be identified. The first issue encountered was related to the language processing. In the interest of evaluating the developed algorithm, we tested it separately on a set of 100 randomly selected images while closely monitoring each step of the execution. Out of the 700 hashtags detected (492 English and 208 German), 208 were segmented incorrectly. Over 50% of the incorrect segmentation occurred due to mixed languages usage and, consequently, erroneous language identification. Furthermore, the algorithm identified in total 857 words (656 English and 201 German words) that could represent an affective state or quality. Only 462 words (or their corresponding synonyms or hypernyms) matched an entry in one of the affective lexicons. In total, 40 words were German words misclassified as English and 64 words were English words misclassified as German. Thus, the implications of the erroneous language processing on the extraction of affective perceptions from the metadata of geotagged photographs were twofold. Due to incorrect language detection, on one hand, the hashtags segmentation algorithm performed erroneously. On the other hand, the computation of the valence and arousal scores was not possible. The improvement of language detection is a challenging task due to the multilingual nature of texts. Nevertheless, to maximize the performance of the sentiment analysis, these issues need to be further studied and better language detection strategies should be applied, which will be part of the future work planned.

Moreover, the proposed approach only considers metadata entries that refer to specific places situated in the study area, and it assumes that, as long as the content's metadata includes a toponym, affective responses towards that place or its affective qualities are expressed. However, this assumption is not sustained by empirical evidence and represents an expedient approach to the matter. A possible improvement of the methodology concerning this aspect is represented by the employment of topic modelling tools to discover the topics that occur in the LBSM content. This is regarded as future work as well.

With reference to the aggregation of the extracted affective ratings, two possible limitations of the approach need to be addressed. First, the proposed approach assumes that affective responses from a large number of users can be aggregated (averaged) to approximate the collective affective evaluation of the area surrounding one specific street segment. Notwithstanding, people's affective responses to environments are subjective in nature and abstract values. Consequently, it is unclear to what extent averaging abstract subjective data is suitable for the identification of collective affective perceptions. This aspect should be subject to further investigations. Secondly, to enable the implementation of the approach and following the idea of a perceptual spatial extent, the affective responses located within a radius of 100 m surrounding each street segment were considered as influential in the calculation of the valence and arousal values of the collective affective ratings. Nevertheless, this aspect needs to be revised since the choice is not based on empirical evidence and might lead to further data validity issues. In addition to these limitations of the study, utilizing LBSM data in research requires constant contemplation upon further data quality issues, such as positional accuracy, or privacy concerns.

To conclude, the conducted research lays out the main issues to be considered when working with LBSM data and presents a fairly solid methodological framework to extract affective information from textual metadata of images shared on Instagram and Flickr. The study also provides new lines of evidence regarding the selectivity bias in terms of places' popularity and the general tendency of social media users to share content with a rather positive affective connotation. Moreover, despite the numerous quality issues identified, on account of the great number of affective perceptions captured, it can be asserted that LBSM data represent a significant source to be regarded in the collection of affective information. We suggest that, for the development of human-centred route planning algorithms, LBSM data could be considered as a weighty additional source to fill in the gaps in the traditionally acquired or crowd-sourced data sets.

Notes

1. http://openstreetmap.org
2. http://www.twitter.com
3. http://www.instagram.com
4. http://www.flickr.com
5. In the interest of filtering out the entries that are not related to places or perceived environments, a rudimentary version of a gazetteer was compiled by creating a list of the toponyms (both in English and German) associated with the study area. The gazetteer (a collection of street and district names, points of interest, land use categories, amenities, and places of worship) containing 18,863 entries was compiled gathering data from OpenStreetMap, Geonames, and Wikitravel.
6. https://pypi.org/project/polyglot/
7. https://pypi.org/project/pyspellchecker/
8. The English collection of words contains 127,156 entries and was assembled by combining the WordNet (https://wordnet.princeton.edu) lexical database (a database of English nouns, verbs, adjectives, and adverbs that are grouped into sets of cognitive synonyms) with the gazetteer compiled for placename identification.
9. For the compilation of the German word list, the same gazetteer was merged with a list of German words with slightly more than 1.9 million entries available online and created by Jan Schreiber in 2017 (https://sourceforge.net/projects/germandict).

Acknowledgements

We would like to thank Dr.-Ing. A Dunkel for providing the Flickr and Instagram data necessary for the implementation of the methodology developed.

ORCID

Eva Hauthal https://orcid.org/0000-0001-8917-600X
Dirk Burghardt https://orcid.org/0000-0003-2949-4887

References

Bao, Siya; Nitta, Tomoyuki; Yanagisawa, Masao; and Togawa, Nozomu: *A safe and comprehensive route finding algorithm for pedestrians based on lighting and landmark conditions*. IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, E100.A(11), 2017, 2439–2450. doi: 10.1587/transfun.E100.A.2439

Bondi, Liz; Davidson, Joyce; and Smith, Mick: *Introduction: geography's 'emotional turn'*. In: Davidson, Joyce; Bondi, Liz; and Smith, Mick (eds.), *Emotional Geographies*, London, UK: Routledge, 2006. 1–16. doi: 10.4324/9781315579245

Haque, Shazia; Kulik, Lars; and Klippel, Alexander: *Algorithms for reliable navigation and wayfinding*. Proceedings of the 2006 International Conference on Spatial Cognition V: Reasoning, Action, Interaction, 2007, 308–326

Hauthal, Eva and Burghardt, Dirk: *Mapping space-related emotions out of user-generated photo metadata considering grammatical issues*. The Cartographic Journal, 53(1), 2016, 78–90. doi: 10.1179/1743277414Y.0000000094

Hollenstein, Livia and Purves, Ross: *Exploring place through user-generated content: using Flickr tags to describe city cores*. Journal of Spatial Information Science, 1, 2010, 21–48. doi: 10.5311/JOSIS.2010.1.3

Huang, Haosheng; Klettner, Silvia; Schmidt, Manuela; et al.: *Affectroute–considering people's affective responses to environments for enhancing route-planning services*. International Journal of Geographical Information Science, 28(12), 2014, 2456–2473. doi: 10.1080/13658816.2014.931585

Huang, Hoasheng: *Context-aware location recommendation using geotagged photos in social media*. ISPRS International Journal of Geo-Information, 5(11), 2016, 195. doi: 10.3390/ijgi5110195

Huang, Hoasheng; Gartner, Georg; and Turdean, Timea: *Social media data as a source for studying people's perception and knowledge of environments*. Mitteilungen der Österreichischen Geographischen Gesellschaft, 155, 2013, 291–302. doi: 10.1553/moegg155s291

Jankowski, Piotr; Andrienko, Natalia; Andrienko, Gennady; and Kisilevich, Slava: *Discovering landmark preferences and movement patterns from photo postings*. Transactions in GIS, 14(6), 2010, 833–852. doi: 10.1111/j.1467-9671.2010.01235.x

Mehrabian, Albert and Russell, James A: *An approach to environmental psychology*. Cambridge, MA, USA: MIT Press, 1974

Miura, Hirokazu; Takeshima, Syujo; Matsuda, Noriyuki; and Taki, Hirokazu: *A study on navigation system for pedestrians based on street illuminations*. 2011, 49–55. doi: 10.1007/978-3-642-23854-3_6

Mody, Ruturaj N; Willis, Katharine S; and Kerstein, Roland: *WiMo: location-based emotion tagging*. 2009, 14. doi: 10.1145/1658550.1658564

Novack, Tessio; Wang, Zhiyong; and Zipf, Alexander: *A system for generating customized pleasant pedestrian routes based on OpenStreetMap data*. Sensors, 18(11), 2018, 3794. doi: 10.3390/s18113794

Pippig, Karsten; Burghardt, Dirk; and Pretchtel, Nikolas: *Semantic similarity analysis of user-generated content for theme-based route planning*. Journal of Location Based Services, 7(4), 2013, 223–245. doi: 10.1080/17489725.2013.804214

Russell, James A: *Core affect and the psychological construction of emotion*. Psychological Review, 110(1), 2003, 145–172. doi: 10.1037/0033-295X.110.1.145

Russell, James A and Pratt, Geraldine: *A description of the affective quality attributed to environments*. Journal of Personality and Social Psychology, 38(2), 1980, 311–322. doi: 10.1037/0022-3514.38.2.311

Ulrich, Roger S: *Aestethic and affective response to natural environment*. In: Altman, Irwin and Wohlwill, Joachim F (eds.), *Behavior and the Natural Environment*, Boston, MA, USA: Springer, 1983. 85–125. doi: 10.1007/978-1-4613-3539-9_4

Võ, Melissa LH; Conrad, Markus; Kuchinke, Lars; et al.: *The Berlin affective word list reloaded (BAWL-R)*. Behavior Research Methods, 41(2), 2009, 534–538. doi: 10.3758/BRM.41.2.534

Warriner, Amy Beth; Kuperman, Victor; and Brysbaert, Marc: *Norms of valence, arousal, and dominance for 13,915 English lemmas*. Behavior Research Methods, 45(4), 2013, 1191–1207. doi: 10.3758/s13428-012-0314-x

Winter, Stephan: *Modeling costs of turns in route planning*. Geoinformatica, 6(4), 2002, 345–361. doi: 10.1023/A:1020853410145

Yik, Michelle; Russell, James A; and Steiger, James H: *A 12-point circumplex structure of core affect*. Emotion, 11(4), 2011, 705–731. doi: 10.1037/a0023980

Geotagging Matters? The Interplay of Space and Place in Politicized Online Social Media Networks

Adrian Tear

School of the Environment, Geography and Geosciences, University of Portsmouth, UK

Voting results in marginal constituencies often determine wider political outcomes. It is now apparent that key electorates in these areas have been geo-behaviourally targeted by elaborate operations intended to manipulate results through advertising, (mis)information, and/or 'fake news' disseminated via online social networks. Attempts to track the geographical diffusion of cyber politicking are hindered by incomplete geospatial referencing in social media (meta)data. Just about 1–2% of publicly posted Twitter tweets, and even fewer Facebook posts, are typically 'geotagged' with Latitude and Longitude coordinates. Many more records (about 25%) make toponymic mention of place. This paper examines about 8 million social media interactions, over 350,000 of which are geotagged, created during the 2012 US Presidential Election and the 2014 Scottish Independence Referendum campaigns, to assess the interplay of space and place in online communications. Results of text and data-mining show that coordinate-geotagging users of Twitter and Facebook, (a) make fewer references to place in their message text, (b) link to articles making fewer mentions of place in their content, and (c) make far fewer links to external content than their non-coordinate-geotagging peers. Despite providing some valuable geospatial information, coordinate-geotagged interactions offer only an inadequate proxy for tracking the spread of all places, linked content, or (mis)information shared online. As Twitter retires its tweet spatialization functionality, new regulatory and technical responses together with a better understanding of place will be required if electoral officials, platform operators, and researchers are to more easily and accurately identify nefarious content targeting specific areas as well as specific individuals during democratic elections.

Keywords: social media; geotagging; natural language processing; place detection

History: received on 7 October 2019; accepted on 27 November 2019; published on 27 January 2020

1 Introduction

Online social networks, politics, and Big Data are in the news. Alarming revelations surrounding Cambridge Analytica's misuse of data for political marketing purposes prompted a US Congressional Committee to investigate data usage, sharing, and privacy policies at Facebook (U.S. House of Representatives, 2018), eventually leading to the imposition of a 'record-breaking' $5 billion fine from the US Federal Trade Commission (Shepardson, 2019). Political campaigning using advanced behavioural and psychographic targeting, alongside geographical micro-marketing designed to bring out or win over key voters (Albright, 2017), may even have affected the outcome of the 2016 US Presidential Election, a contest which Cambridge Analytica claimed to have 'won' for Donald Trump (Lewis and Hilder, 2018).

Steiger et al. (2015, p. 816) note that coordinate-geotagged online social network (OSN) interactions, sourced primarily from Twitter, have demonstrated high degrees of utility in 'research on event detection

A Tear (2020): *Geotagging Matters? The Interplay of Space and Place in Politicized Online Social Media Networks*. In: FB Mocnik and R Westerholt (eds.), Proceedings of the 2nd International Symposium on Platial Information Science (PLATIAL'19), pp. 61–72

https://doi.org/10.5281/zenodo.3628879

PLATIAL'19 Second International Symposium on Platial Information Science (PLATIAL'19)
Coventry, UK; 5–6 September 2019

[particularly in the] investigation of abnormal spatial, temporal and semantic tweet frequencies [surrounding] disaster and emergency [situations]'. The current research uses a mixture of coordinate-geotagged and non-coordinate-geotagged social media data from Facebook and Twitter, collected during the 2012 US Presidential Election (US2012) and the 2014 Scottish Independence Referendum (SCOT2014), to determine whether a similarly high level of analytical utility may be observed in political contexts. Understanding how different classes of social media users imprint their communications with place or, less frequently, space – or consume, link to, and share third party Uniform Resource Locator (URL) content imprinted with place – is essential when attempting to accurately track the downstream diffusion of deliberately geo-targeted political advertising.

2 Geographical Characteristics of Social Media Data

2.1 Space and Place

Senses of known-place(s), affirmed-place(s), and space(s), some of which may be accompanied by apparently accurate latitude and longitude coordinates, are often highly conflated in social media data. Users of Twitter, e.g., when registering, are asked 'Where in the world are you?' (Hecht et al., 2011) and may just as reasonably answer 'BRICK city bitch' or 'Somewhere, Overthere' as 'Concord, NC' or 'iPhone: 40.699490,-73.891556'. Difficulties inherent in identifying and parsing Potential Geographic Information (PGI) in free-form social media message text and associated (meta)data are amplified considerably when, as in this research, place-based geographical references must be detected computationally. Consequently, and as huge data volumes preclude individual human examination of over 8 million social media interactions, necessarily focused definitions of 'space' and 'place' are adopted in this research:

- **Space** - refers to geographically and explicitly *locational* data, i.e., to a point defined by a pair of latitude and longitude coordinates.
- **Place** – refers to computationally-identifiable geographical references in text, i.e., to *toponymic* place names, e.g., of towns, cities, counties, states, countries, etc.

Space, where it exists in social media interaction (meta)data, may generally be regarded unambiguously; the latitude and longitude coordinates of a user's location have been recorded alongside their message text by a Global Positioning System (GPS) equipped mobile device just at the moment of message creation (Li et al., 2016). Place, in social media data, retains many of the elements of ambiguity identified by Tuan (1977) and other geographical theorists but is referenced, on the admittedly narrower grounds adopted here, much more widely in message text, metadata, and linked/shared content than space.

2.2 Geotagging Rates and Behaviour

Typically, and somewhat unfortunately for geographers, only small percentages of social media interactions are geotagged with latitude and longitude coordinates. Leetaru et al. (2013) report that just 1.6% of about 1.5 billion Twitter interactions analysed in their study contained 'Exact locations'. Slightly higher rates have been reported elsewhere (Croitoru et al., 2013) with variability attributed to event type (e.g., an elevated 16% following the Fukushima nuclear disaster in Japan), cultural practice (e.g., some nations use OSNs more frequently than others), and differing technological factors (e.g., smartphone adoption rates). While most users' mobile devices are perfectly capable of imprinting coordinates alongside their OSN posts, geotagging is an 'opt-in' feature which users must explicitly enable in their software application (Sui, 2017). Few users choose to deliberately activate geotagging facilities (Tasse et al., 2017) and, consequently, most mapping and geographical analyses of social media interactions are enabled not by the majority of OSN users, but by a *distinct minority* who choose to post with coordinates.

The current research questions whether there is an over-reliance on 'geosocial' data deposited by just about 1–2% of all social media users and whether expressions of 'place' in message text and linked/shared content are highly correlated with 'space' in coordinate-geotagged OSN interactions.

The work addresses a fundamental question: who makes, or links to external content containing, the most place-based references on social media networks, i.e., who is among the coordinate-geotagging or non-coordinate-geotagging users of these sites? Answering this question, to determine the toponymical representativeness of coordinate-geotagging users, helps determine whether or not this minority group may be used as 'markers' to accurately and spatially, through their latitude and longitude coordinates, trace the geographical diffusion of online opinion, and/or (mis)information.

3 Data Subjects

The subjects of this research are politically discursive social media messages; 8,196,380 OSN interactions created by 2,436,167 individual users of Twitter and Facebook in a roughly 90:10 ratio during two case study electoral events. Four sampled 'streams' of social media data were collected using the DataSift platform, a content aggregation system capable (at the time) of accessing Twitter's full 'Firehose' of tweets, together with messages publicly-posted on Facebook. In each case, the source files in both Comma Separated Values (CSV) and JavaScript Object Notation (JSON) formats (ECMA International, 2017) were loaded into the Oracle 12c Relational Database Management System (RDBMS). The data collected for each event are further described below.

3.1 2012 US Presidential Election

During the two-month run up to the US Presidential Election of 6 November 2012, 1,661,402 Twitter tweets and 57,265 Facebook posts were sampled from contemporaneous OSN communications. Three sample sets from Twitter and Facebook were recorded, filtered on a range of identical text search terms, and controlled for explicit presence/absence of geographical coordinates, extent (country), and/or language. The interactions were filtered on any case-insensitive words or phrases matching those illustrated in Figure 1, which shows the contribution of each search term to the sample, noting that some terms (e.g., 'US President' and 'Obama') may have appeared multiple times within message text. Despite filtering on 15 search terms the top 3 terms account for 78.97% of the interactions sampled in the data set. Filtering selected for inclusion mainly on candidate surname, forename, or a combination of the two. In both political events (see also Figure 2) the top two terms usefully, and reasonably evenly, select interactions for the major protagonists in both contests. The three US2012 streams were recorded, stored, and downloaded from DataSift's servers. The data set consists of 1,718,667 rows across three files each with up to 146 fields.

3.2 2014 Scottish Independence Referendum

The necessity for a second case study was prompted by early analysis of data from the first. The US2012 data set consisted of three sampled streams. Sampling was used to restrict data volumes and control costs, but also resulted in an 'incomplete' data set where the full network graph of tweeting, mentioning, and retweeting could not be examined. Consequently, a second real-time recording of OSN interactions during the much longer run up to the 2014 Scottish Independence Referendum was started on 18 September 2013, exactly one year before the vote was due to take place.

Scotland, with a much smaller population (about 5 million) than the US (about 320 million), was thought unlikely to create the sorts of OSN data volumes a 1:1 sampled US recording would have generated. Interactions were filtered on any case-insensitive words or phrases matching those illustrated in Figure 2, which also shows the contribution of each search term to the sample, again noting that some terms will have appeared multiple times within message text. Deliberate misspellings ('independance', etc.) were incorporated in the filter design as misspellings are a common feature of OSN communications (Deitrick and Hu, 2013).

Despite Scotland's small population size, worldwide interest in the outcome of the referendum, coupled with the longer-running nature of the recording, eventually resulted in the collection of about 6.5 million OSN interactions. The top 3 of 27 search terms account for 63.25% of interactions sampled in the data set. Filtering has selected for inclusion on a mix of First Minister (and Vote Yes leader) Alex Salmond's surname, the campaign slogan ('Better Together') of the Vote No (remain united) coalition, where no one political figure spearheaded the campaign, and the abbreviation 'SNP' (Scottish

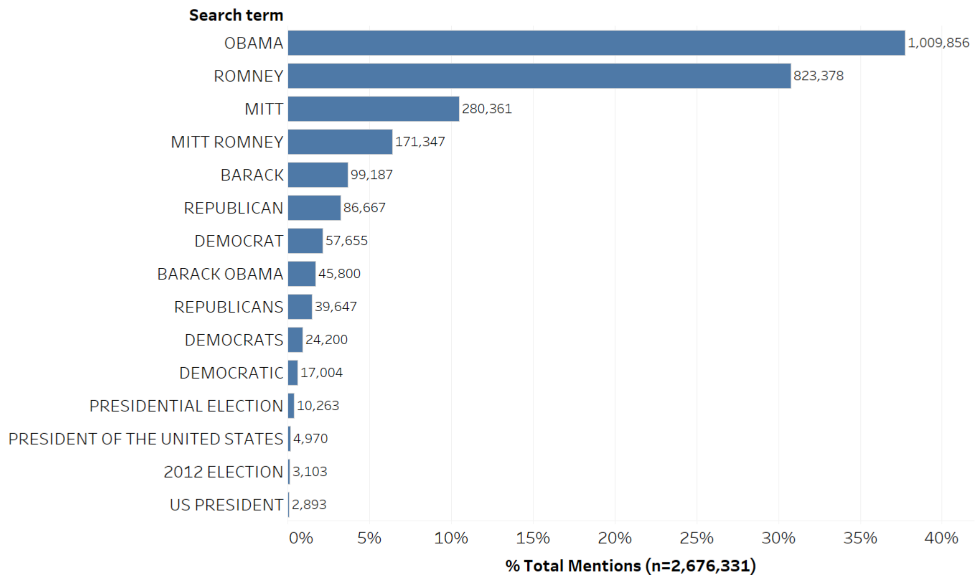


Figure 1: US2012: Search terms used. Numeric and percentage contribution to OSN interactions sampled (2,676,331 total mentions of search terms in 1,718,667 messages).

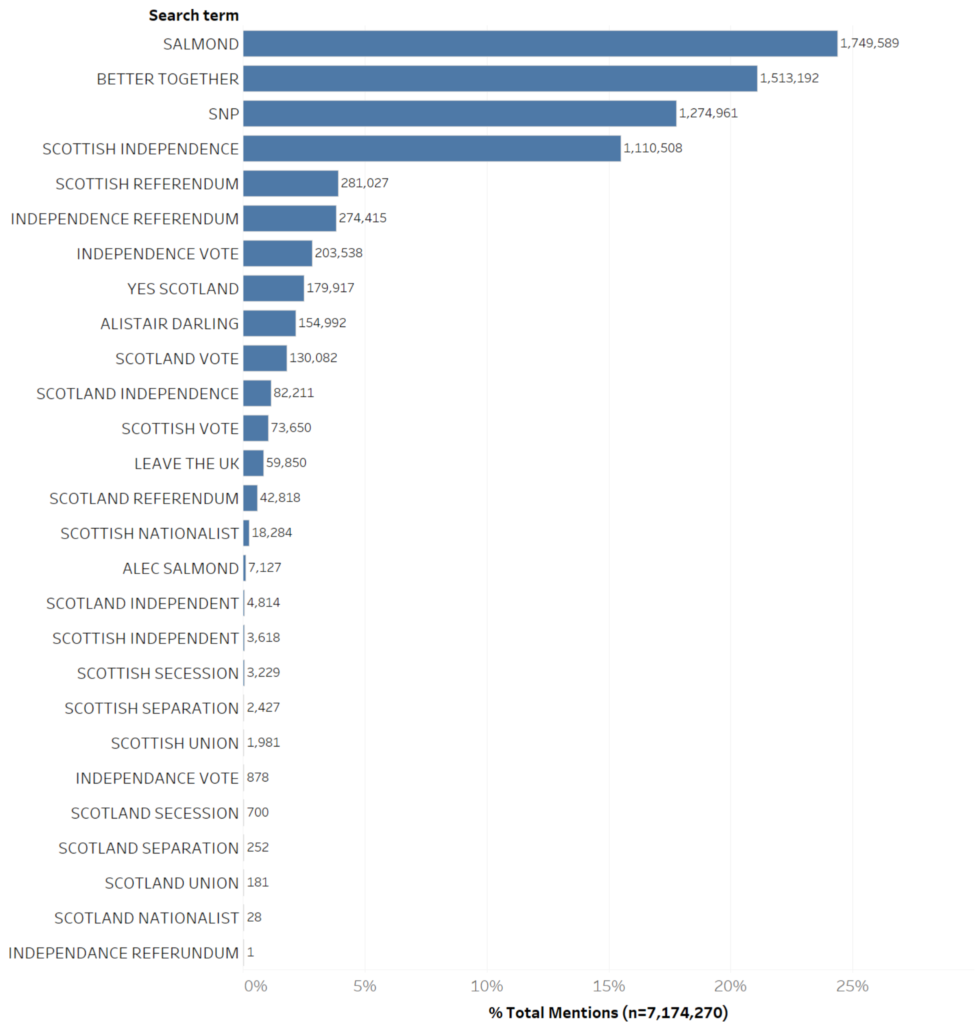


Figure 2: SCOT2014: Search terms used. Numeric and percentage contribution to OSN interactions sampled (7,174,270 total mentions of search terms in 6,477,713 messages).

Nationalist Party), the name of the pro-independence party in Scotland. As in the US2012 event, the top two terms usefully, and reasonably evenly, select interactions for the major protagonists in the 2014 Scottish Independence Referendum and are thought to offer a balance of messages for inclusion in the sample for both opposing sides of the political debate.

4 Data Analysis

Stock (2018, p. 209) has noted that 'During the last ten years, a large body of research extracting and analysing geographic data from social media has developed'. Reviewing 690 papers accessing 20 social media platforms she states that 'a wide array of [...] approaches have been developed, with methods that extract place names from message text providing the highest accuracy'. The three NLP packages used in this research to detect toponymic mentions in message text and, additionally, in linked/shared URL content are summarized below.

- **GATE** – The General Architecture for Text Engineering's TwitIE processor, a Twitter Information Extraction engine, and 'open-source NLP pipeline customized to microblog text at every stage' (Bontcheva et al., 2013), running on the highly-scaleable GATEcloud.net system (Tablan et al., 2013), was used to identify various entities, e.g., toponyms, people, or organizations, mentioned in message text. GATE cannot, yet, append coordinates to detected platial references.
- **CLAVIN** – The Cartographic Location and Vicinity Indexer 'extracts location names from unstructured text and resolves them against the GeoNames gazetteer to produce data-rich geographic entities' (Berico-Technologies, 2017). As a specialist 'geo-parser', CLAVIN appends GeoNames-derived coordinate pairs to toponymic place names detected in interaction message text. CLAVIN cannot, however, identify other entity types (e.g., people, organizations, etc.) in text.
- **AlchemyAPI** – AlchemyAPI, now re-branded as Watson Natural Language Understanding (IBM, 2017), is Cloud-hosted, commercial software. Academic usage is restricted to 30,000 'daily transactions', a processing restriction which prevented the timely processing of all about 8 million OSN interactions. Instead, a sample of 311,575 messages were processed, alongside 641,472 distinct URLs shared 3,485,840 times in the research data corpus, just 45,492 of which were shared by coordinate-geotagging users.

All about 8 million OSN messages were passed through TwitIE on GATEcloud and CLAVIN-rest. AlchemyAPI was used to detect toponyms in the linked/shared content of OSN users. The analytical pipeline, centred around an Oracle 12c database, relied upon both Cloud-hosted and locally virtualized computing resources running CentOS or Ubuntu Linux under Oracle VirtualBox on a Windows 10 host.

5 Results

Despite adopting technically different solutions to the 'challenge' of toponymic place detection in free form text (Li et al., 2016) both GATEcloud and CLAVIN-rest, the two systems successfully used to text-mine all about 8 million social media messages in the research data corpus, produced highly comparable results.

Table 1 shows the number of locations resolved by GATEcloud in interaction message text. The main entity type of interest returned in GATEcloud JSON following TwitIE processing (Bontcheva et al., 2013) is `Location`, in the `locType` key, which references indexed characters (i.e., the n-th characters in the message text) containing the detected location. These are coded as `region`, `province`, `post`, `unknown`, `country`, `country_abbrev`, `city`, `airport`, `racecourse`, or `pre`. These codings are mainly self-evident, except for `pre` and `post` which refer to first/last matches to parts of a location such as 'Mount', 'East', 'Cape', 'Isle of', etc., which co-occur with a proper noun. TwitIE on GATEcloud detects `locType` in the message text of 263,296 (15.32% of all) US2012 interactions, identifying a further 2,088,788 messages containing locations (32.25% of all) in the SCOT2014 data set. The ratio of resolved locations per interaction is higher in the larger and more recent SCOT2014 data set and is significantly higher for message text sourced from Facebook.

Table 1: GATEcloud place detection results. US2012/SCOT2014: Number of resolved locations detected by GATEcloud in Facebook (FB), Twitter tweet (TW), and Twitter retweet (RT) interactions.

	US2012	Ratio	SCOT2014	Ratio	Total
FB resolved locations	25,405	1.90	2,097,506	5.66	2,122,911
FB n interactions	13,341		370,774		384,115
TW resolved locations	153,085	1.23	1,087,698	1.31	1,240,783
TW n interactions	123,960		833,235		957,195
RT resolved locations	151,899	1.21	1,130,344	1.28	1,282,243
RT n interactions	125,995		884,779		1,010,774
Total Resolved	330,389	1.25	4,315,548	2.07	4,645,937
Total Interactions	263,296		2,088,788		2,352,084

Table 2: CLAVIN-rest place detection results. US2012/SCOT2014: Number of resolved locations detected by CLAVIN-rest in Facebook (FB), Twitter tweet (TW), and Twitter retweet (RT) interactions.

	US2012	Ratio	SCOT2014	Ratio	Total
FB resolved locations	22,019	1.80	1,573,145	3.98	1,595,166
FB n interactions	12,199		395,112		407,311
TW resolved locations	120,491	1.19	832,941	1.25	953,433
TW n interactions	101,664		667,263		768,927
RT resolved locations	119,979	1.18	856,383	1.22	976,363
RT n interactions	101,599		700,567		802,166
Total Resolved	262,489	1.22	3,262,469	1.85	3,524,959
Total Interactions	215,462		1,762,942		1,978,404

Table 2 shows the number of locations resolved by CLAVIN-rest in interaction message text. It is apparent, as with TwitIE on GATEcloud, that 45.25% of resolved locations (n = 1,595,166) detected by CLAVIN-rest stem from 407,311 geoparsed OSN interactions sourced from Facebook (i.e., 20.59% of all interactions processed), most of which (n = 395,112) were collected during the 2014 Scottish Independence Referendum. Larger numbers of Twitter tweet (n = 768,927) and retweet (n = 802,166) interactions (total n = 1,571,093) were geoparsed by CLAVIN-rest but yielded a total of only 1,929,794 (n = 953,433 and n = 976,363, respectively) resolved locations, 87.54% of which (n = 1,689,324) were found in the SCOT2014 data set which features a higher proportion of Twitter retweets. Most locations resolved in retweets will, of course, be duplicates of locations found in the originating tweet.

Welch Two Sample T-tests, calculated using R (The R Foundation, 2018), compare distributions of numbers of NLP-detected toponymic mentions per interaction, or per user, for non-coordinate-geotagged/ing and coordinate-geotagged/ing interactions/users (Table 3). During the US2012 event, in 9 out of 20 cases like-for-like comparisons of geoparser, OSN source and level for non-coordinate-geotagged message text or linked/shared URL content against coordinate-geotagged corollaries are statistically significant with >95% confidence. The null hypothesis, that there is no difference in the distribution of numbers of toponymic mentions detected in OSN message text or linked/shared URL content by OSN source for the given NLP/geoparsers at interaction and user levels, can be rejected. In most (n = 6) of these statistically significant comparisons non-coordinate-geotagged interactions or non-coordinate-geotagging users make more NLP/geoparser-detectable toponymic mentions in message text, or link to and share URLs having more detectable toponymic mentions in content, than their coordinate-geotagged or geotagging corollaries. Statistics for 6 cases (marked 'N/A' in Table 3) cannot be calculated in R for the US2012 data set as no coordinate-geotagged Facebook interactions are present. During the SCOT2014 event, statistical significance with >95% confidence is found in 18 of 20 like-for-like cases comparing numbers of toponymic detections by NLP/geoparser in message text and linked/shared URLs for Facebook, Twitter tweet, and retweet data at interaction and user levels. In over half (n = 11) of these statistically significant comparisons non-coordinate-geotagged interactions or non-coordinate-geotagging users make more NLP/geoparser-detectable toponymic mentions in message text, or link to and share URLs having more detectable toponymic mentions in content, than their

Table 3: Statistical results. Welch Two Sample T-tests, at interaction and user levels, comparing the number of resolved locations detected by the three NLP systems in coordinate-geotagged and non-coordinate-geotagged Facebook (FB), Twitter tweet (TW), and Twitter retweet (RT) message text and linked/shared content.

System	Level	Source	US2012		SCOT2014	
			t	$t > \pm 2$	t	$t > \pm 2$
GATEcloud (Messages)	Interaction	FB	N/A	N/A	1.03	✗
GATEcloud (Messages)	Interaction	TW	-32.54***	✓	3.37***	✓
GATEcloud (Messages)	Interaction	RT	0.44	✗	7.01***	✓
AlchemyAPI (Messages)	Interaction	TW	-1.04	✗	-3.55***	✓
CLAVIN-rest (Messages)	Interaction	FB	N/A	N/A	4.91***	✓
CLAVIN-rest (Messages)	Interaction	TW	3.92***	✓	2.99**	✓
CLAVIN-rest (Messages)	Interaction	RT	2.63**	✓	4.68***	✓
AlchemyAPI (Links)	Interaction	FB	N/A	N/A	1.10	✗
AlchemyAPI (Links)	Interaction	TW	1.29	✗	-29.85***	✓
AlchemyAPI (Links)	Interaction	RT	2.21*	✓	3.44***	✓
GATEcloud (Messages)	User	FB	N/A	N/A	83.69***	✓
GATEcloud (Messages)	User	TW	-34.48***	✓	69.32***	✓
GATEcloud (Messages)	User	RT	14.01***	✓	-13.52***	✓
AlchemyAPI (Messages)	User	TW	-20.24***	✓	-7.17***	✓
CLAVIN-rest (Messages)	User	FB	N/A	N/A	97.36***	✓
CLAVIN-rest (Messages)	User	TW	37.18***	✓	47.42***	✓
CLAVIN-rest (Messages)	User	RT	1.25	✗	-12.41***	✓
AlchemyAPI (Links)	User	FB	N/A	N/A	52.12***	✓
AlchemyAPI (Links)	User	TW	2.92**	✓	-33.01***	✓
AlchemyAPI (Links)	User	RT	0.25	✗	-7.20***	✓

* $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.001$

coordinate-geotagged or coordinate-geotagging corollaries.

These findings are at odds with the research hypothesis that coordinate-geotagging users are the most geographically expressive of all OSN users. Although they do actively (or accidentally) coordinate-geotag their Twitter tweets or Facebook posts, this small group of social media users are not, in three important respects, representative of all OSN users. Of course, OSN users in general (Diaz et al., 2016), and geotagging users in particular (Sloan and Morgan, 2015), are not thought to be representative of the general population. During elections, they are likely to be even less so, probably being younger and living in urban areas and often, according to Barberá and Rivero (2015), exhibiting 'extreme ideological preferences'. The results of a comprehensive analysis and cross-comparison of toponymic mentions detected in the message text and URL link shares of coordinate-geotagging and non-coordinate-geotagging users interacting online during two data-rich political case study events show that

1. coordinate-geotagging users make fewer toponymic mentions in message text than non-coordinate-geotagging users of two popular OSN platforms;

2. coordinate-geotagging users make far fewer URL link shares than non-coordinate-geotagging users, and

3. the content of URLs shared by coordinate-geotagging users makes fewer mentions of place than content shared by non-coordinate-geotagging users.

The research findings presented here imply that geographical outputs (such as point maps and counts or aggregations to larger areal units such as constituencies or states) based on searches for specific words, toponyms, #hashtags or @mentions in message text or URL link shares, which may readily be mapped using the interaction latitude and longitude coordinates of Twitter tweets or Facebook posts deposited by coordinate-geotagging users (or aggregated to wider areas using a GIS), are unlikely to be representative of the spread of all such content within online social networks.

6 Discussion

Just as Massey and Allen's *Geography Matters!* (1984) reaffirmed the relevance of geography in socio-spatial, environmental, political, and economic spheres, a conception of place clearly matters when individuals interact online using social media platforms. In the 2012 US Presidential Election and the 2014 Scottish Independence Referendum case studies examined here, about 3.5–4.5 million toponymic mentions have been identified in around one quarter of the about 8 million interactions in the research data corpus. Around one quarter of the about 7 million entities identified in about 650,000 distinct URLs – posted, tweeted, or retweeted about 3.5 million times – also contained toponymically identifiable content. Elections are peculiarly *geographic*, as well as *political*, events. It is, therefore, both unsurprising and reassuring to find that electorates and commentators make frequent geographical references online during electoral campaigns, and that many of these mentions refer to the 'swing' states or constituencies the ballot results of which typically shape wider political outcomes.

What then of geotagging; does it *matter*? Geotagging is a relatively recent socio-technological phenomenon, primarily enabled by the worldwide proliferation and usage of GPS-equipped mobile, or smartphone, devices. The increasingly large volumes of Ambient (and/or Volunteered) Geospatial Information now available (Goodchild, 2007; Stefanidis et al., 2013b) offer new research opportunities for scholars in geography and related social science disciplines. Increased scrutiny of 'Geo-social Networks' (Bahir and Peled, 2013), and the possibilities they afford for wider geographical analysis, are demonstrated by the growing number of academic articles and specialist journals published in the last decade or so, many of them cited in this paper. Geotagged photographic images publicly posted on Flickr have been used to combat wildlife poaching in protected areas and in criminological research (Lemieux, 2015). Geotagged social media and other 'Big Data' have been used to monitor natural disaster situations (Burns, 2018; Goodchild and Glennon, 2010). OpenStreetMap has been used in the study of the production and 'prosumption' of user-generated geographic Big Data (Cockayne, 2016). Human interaction data sourced from Twitter and, to a lesser extent, Facebook have been used, seemingly, to 'do everything'; from monitoring earthquakes (Crooks et al., 2013) to tracking riots (Bonilla and Rosa, 2015; Crampton et al., 2013), helping to demarcate urban areas (Yin et al., 2017), and much else besides (see Kapoor et al., 2018, for a recent and comprehensive summary of application areas). This proliferation of research activity, e.g., '[delineating] city cores, [gaining] insights into travel plans and tourism, [characterizing] urban landscapes, [studying] global migrations or [identifying] mobility patterns', has also been identified by Rzeszewski and Beluch (2017), who go on to note that 'much less attention' has been devoted to studies investigating the 'subgroup of users that produce (or rather contribute since they may not be aware of it) [. . .] ambient geospatial information' on social media networks.

7 Conclusion

Bespoke geographical targeting campaigns, as developed by Cambridge Analytica (Albright, 2017), may exploit toponymic references found in users' self-reported 'Location' fields (Hecht et al., 2011), toponymic references found in users' publicly-posted message text (Stock, 2018), and/or latitude and longitude coordinates deposited in OSN metadata when users optionally choose to 'geotag' their social media posts (Kumar et al., 2014). On Facebook (2018), advertisements can be displayed to all users in (or within a radius of) selected locations, users who live in those locations (also 'validated' by Internet Protocol, IP, address), users currently in those locations ('as determined only by mobile device') or people just passing through ('as determined by mobile device [when it is] greater than 100 miles from their stated home location from their Facebook profile'). Other major social media websites operated by Google, Instagram, Snapchat, Twitter, and YouTube offer broadly similar facilities to advertisers – or political campaigners – using their services. All also offer targeting on age, basic demographics (e.g., gender), and, in several cases, on more advanced behavioural or similarity traits (e.g., interests and 'Lookalike Audiences' in Facebook's case). It is currently unclear whether recent attempts to distort the outcome of democratic elections through geo-behavioural targeting have shown clear 'monolithic effects [but] the impact of social media in political campaigning around the world is undeniable' (Dimitrova and Matthes, 2018). It is also, unfortunately for concerned citizens, electoral regulators, and others, much easier to set up a geo-targeted online political advertising campaign than it is for third parties to monitor the downstream geographical diffusion or effectiveness of such communications.

Coordinate-geotagging users are much less widely followed than others on OSNs and, somewhat counter-intuitively, express themselves less geographically than others in their message text and through the URLs they choose to link to and share. The comprehensive analysis of social media interactions presented here reveals important differences in the posting behaviour of coordinate-geotagging and non-coordinate-geotagging users during two political case study events. The more fundamental question, whether geotagging matters, is not so easily answered. To a professional geographer, the vast number of coordinate pairs now deposited online by social media users appears highly propitious. On a massive scale, arguably for the first time in human history, it is possible to know who is saying what, when, and where. However, as detailed earlier, and remarked upon by Paraskevopoulos and Palpanas (2016, p. 1), 'only a very small percentage of [OSN] posts are geotagged, which significantly restricts the applicability and utility of [many] applications'. Low rates of coordinate-geotagging in OSN data, and the unrepresentativeness of coordinate-geotagging users, definitely limit the 'applicability' of any analyses based solely upon geotagging users' message text, metadata, or spatial location in political contexts. When examining politicized communications made on social media networks, or determining how political opinion or (mis)information may be geographically tracked on these platforms, it appears that *geo* matters, but *tagging* matters much less. As Twitter, the most widely studied OSN (Stock, 2018; Tufekci, 2014), retires its tweet geotagging functionality (Leetaru, 2019) and Facebook, one of the world's key outlets for targeted political advertising (Lilleker et al., 2015), restricts data access in an operational environment ripe for regulation (McKinnon and Seetharaman, 2018), the need to efficiently and accurately detect place, in potentially dwindling volumes of social media message text and metadata, seems certain to increase.

Supplementary Information

This paper is an abridged synopsis of Adrian Tear's doctoral thesis (2018), supervised by Professors Richard Healey and Humphrey Southall in the Department of Geography at the University of Portsmouth. The full text of the thesis is available at https://researchportal.port.ac.uk/portal/en/theses/geotagging-matters(90524d65-f62c-4bfd-b201-642a90368f96).html.

Acknowledgements

The results reported here rely upon computational analysis of over 8 million social media messages publicly posted by just under 2.5 million users of Twitter and Facebook. I thank these users, the two online social network platforms which enable(d) data access and DataSift, without whom it would have been impossible to download and analyse so many individual, albeit terse and often ungrammatical, thoughts and opinions. Professor Kalina Bontcheva and colleagues at the University of Sheffield provided invaluable assistance in the use of GATE and GATEcloud natural language processing software. Further technical assistance was provided by staff at IBM Watson, Oracle Corporation, and the University of Portsmouth.

ORCID

Adrian Tear https://orcid.org/0000-0001-5028-2622

References

Albright, Jonathan: *Cambridge Analytica: the geotargeting and emotional data mining scripts*. https://medium.com/tow-center/cambridge-analytica-the-geotargeting-and-emotional-data-mining-scripts-bcc3c428d77f, 2017. Retrieved 27 November 2019

Bahir, Eitan and Peled, Ammatzia: *Identifying and tracking major events using geo-social networks*. Social Science Computer Review, 31(4), 2013, 458–470. doi: 10.1177/0894439313483689

Barberá, Pablo and Rivero, Gonzalo: *Understanding the political representativeness of Twitter users*. Social Science Computer Review, 33(6), 2015, 712–729. doi: 10.1177/0894439314558836

Berico-Technologies: *About clavin*. https://clavin.bericotechnologies.com/about-clavin/, 2017. Retrieved 27 November 2019

Bonilla, Yarimar and Rosa, Jonathan: *#Ferguson: digital protest, hashtag ethnography, and the racial politics of Social Media in the United States*. American Ethnologist, 42(1), 2015, 4–17. doi: 10.1111/amet.12112

Bontcheva, Kalina; Derczynski, Leon; Funk, Adam; et al.: *TwitIE: an open-source information extraction pipeline for microblog text*. Proceedings of the International Conference on Recent Advances in Natural Language Processing, 2013, 83–90

Burns, Ryan: *Datafying disaster: institutional framings of data production following superstorm sandy*. Annals of the American Association of Geographers, 108(2), 2018, 569–578. doi: 10.1080/24694452.2017.1402673

Cockayne, Daniel G: *Affect and value in critical examinations of the production and 'prosumption of Big Data*. Big Data and Society, 3(2), 2016, 1–11. doi: 10.1177/2053951716640566

Crampton, Jeremy W; Graham, Mark; Poorthuis, Ate; et al.: *Beyond the geotag: situating 'big data' and leveraging the potential of the geoweb*. Cartography and Geographic Information Science, 40(2), 2013, 130–139. doi: 10.1080/15230406.2013.777137

Croitoru, Arie; Crooks, Andrew; Radzikowski, Jacek; and Stefanidis, Anthony: *Geosocial gauge: a system prototype for knowledge discovery from Social Media*. International Journal of Geographical Information Science, 27(12), 2013, 2483–2508. doi: 10.1080/13658816.2013.825724

Crooks, Andrew; Croitoru, Arie; Stefanidis, Anthony; and Radzikowski, Jacek: *#Earthquake: Twitter as a distributed sensor system*. Transactions in GIS, 17(1), 2013, 124–147. doi: 10.1111/j.1467-9671.2012.01359.x

Deitrick, William and Hu, Wei: *Mutually enhancing community detection and sentiment analysis on Twitter networks*. Journal of Data Analysis and Information Processing, 1(3), 2013, 19–29. doi: 10.4236/jdaip.2013.13004

Diaz, Fernando; Gamon, Michael; Hofman, Jake M; Kıcıman, Emre; and Rothschild, David: *Online and social media data as an imperfect continuous panel survey*. PLOS ONE, 11(1), 2016, e0145406. doi: 10.1371/journal.pone.0145406

Dimitrova, Daniela V and Matthes, Jörg: *Social Media in political campaigning around the world: theoretical and methodological challenges*. Journalism and Mass Communication Quarterly, 95(2), 2018, 333–342. doi: 10.1177/1077699018770437

ECMA International: *The JSON data interchange syntax*. Tech. rep., 2017. International Standard

Facebook: *Location targeting*. https://en-gb.facebook.com/business/a/location-targeting, 2018. Retrieved 27 November 2019

Goodchild, Michael F: *Citizens as sensors: the world of volunteered geography*. GeoJournal, 69(4), 2007, 211–221. doi: 10.1007/s10708-007-9111-y

Goodchild, Michael F and Glennon, J Alan: *Crowdsourcing geographic information for disaster response: a research frontier*. International Journal of Digital Earth, 3(3), 2010, 231–241. doi: 10.1080/17538941003759255

Hecht, Brent; Hong, Lichan; Suh, Bongwon; and Chi, Ed H: *Tweets from Justin Bieber's heart*. Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems, 2011, 237–246. doi: 10.1145/1978942.1978976

IBM: *Natural language understanding*. https://console.stage1.bluemix.net/catalog/services/natural-language-understanding, 2017. Retrieved 27 November 2019

Kapoor, Kawaljeet Kaur; Tamilmani, Kuttimani; Rana, Nripendra P; et al.: *Advances in social media research: past, present and future*. Information Systems Frontiers, 20(3), 2018, 531–558. doi: 10.1007/s10796-017-9810-y

Kumar, Shamanth; Morstatter, Fred; and Liu, Huan: *Twitter data analytics*. New York, NY, USA: Springer, 2014. doi: 10.1007/978-1-4614-9372-3

Leetaru, Kalev: *Twitter's shift away from GPS-tagged tweets suggests it will be harder to fight 'deep fakes'*. Newspaper article in Forbes Magazine, 2019

Leetaru, Kalev; Wang, Shaowen; Cao, Guofeng; Padmanabhan, Anand; and Shook, Eric: *Mapping the global Twitter heartbeat: the geography of Twitter*. First Monday, 18(5), 2013

Lemieux, Andrew M: *Geotagged photos: a useful tool for criminological research?* Crime Science, 4, 2015, 3. doi: 10.1186/s40163-015-0017-6

Lewis, Paul and Hilder, Paul: *Leaked: Cambridge Analytica's blueprint for Trump victory*. Newspaper article in The Guardian, 2018

Li, Songnian; Dragicevic, Suzana; Castro, Francesc Anton; et al.: *Geospatial big data handling theory and methods: a review and research challenges*. ISPRS Journal of Photogrammetry and Remote Sensing, 115, 2016, 119–133. doi: 10.1016/j.isprsjprs.2015.10.012

Lilleker, Darren G; Tenscher, Jens; and Štětka, Václav: *Towards hypermedia campaigning? Perceptions of new media's importance for campaigning by party strategists in comparative perspective*. Information, Communication and Society, 18(7), 2015, 747–765. doi: 10.1080/1369118X.2014.993679

Massey, Doreen and Allen, John: *Geography matters!* Cambridge, UK: Cambridge University Press, 1984

McKinnon, John D and Seetharaman, Deepa: *In Facebook hearings, lawmakers ramp up talk of regulation*. Newspaper article in The Wall Street Journal, 2018

Paraskevopoulos, Pavlos and Palpanas, Themis: *Where has this tweet come from? fast and fine-grained geolocalization of non-geotagged tweets*. Social Network Analysis and Mining, 6, 2016, 89. doi: 10.1007/s13278-016-0400-7

Rzeszewski, Michal and Beluch, Lukasz: *Spatial characteristics of Twitter users–toward the understanding of geosocial media production*. ISPRS International Journal of Geo-Information, 6(8), 2017, 236. doi: 10.3390/ijgi6080236

Shepardson, David: *Facebook to pay record $5 billion U.S. fine over privacy; faces antitrust probe*. https://www.reuters.com/article/us-facebook-ftc/facebook-to-pay-record-5-billion-us-fine-over-privacy-faces-antitrust-probe-idUSKCN1UJ1L9, 2019. Retrieved 27 November 2019

Sloan, Luke and Morgan, Jeffrey: *Who tweets with their location? Understanding the relationship between demographic characteristics and the use of geoservices and geotagging on Twitter*. PLoS ONE, 10(11), 2015, e0142209. doi: 10.1371/journal.pone.0142209

Stefanidis, Anthony; Crooks, Andrew; and Radzikowski, Jacek: *Harvesting ambient geospatial information from social media feeds*. GeoJournal, 78(2), 2013b, 319–338. doi: 10.1007/s10708-011-9438-2

Steiger, Enrico; de Albuquerque, João Porto; and Zipf, Alexander: *An advanced systematic literature review on spatiotemporal analyses of Twitter data*. Transactions in GIS, 19(6), 2015, 809–834. doi: 10.1111/tgis.12132

Stock, Kristin: *Mining location from social media: a systematic review*. Computers, Environment and Urban Systems, 71, 2018, 209–240. doi: 10.1016/j.compenvurbsys.2018.05.007

Sui, Daniel Z: *Understanding locational-based services: core technologies, key applications and major concerns*. In: Warf, Barney (ed.), *Handbook on Geographies of Technology*, Cheltenham, UK: Edward Elgar Publishing, 2017. 85–95

Tablan, Valentin; Roberts, Ian; Cunningham, Hamish; and Bontcheva, Kalina: *GATECloud.net: a platform for large-scale, open-source text processing on the cloud*. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 371(1983), 2013, 20120071. doi: 10.1098/rsta.2012.0071

Tasse, Dan; Liu, Zichen; Sciuto, Alex; and Hong, Jason I: *State of the geotags: Motivations and recent changes*. Proceedings of the 11th International AAAI Conference on Web and Social Media, 2017, 250–259

Tear, Adrian: *Geotagging matters? The role of space and place in politicised social media discourse*. Ph.D. thesis, University of Portsmouth, 2018

The R Foundation: *R: the R project for statistical computing*. https://www.r-project.org/, 2018. Retrieved 27 November 2019

Tuan, Yi-Fu: *Space and place. The perspective of experience*. Minneapolis, MN, USA: University of Minnesota Press, 1977

Tufekci, Zeynep: *Big questions for Social Media Big Data: representativeness, validity and other methodological pitfalls*. Proceedings of the 8th International AAAI Conference on Weblogs and Social Media, 2014, 505–514

U.S. House of Representatives: *Facebook: transparency and use of consumer data*. https://energycommerce.house.gov/committee-activity/hearings/hearing-on-facebook-transparency-and-use-of-consumer-data-full-committee, 2018. Retrieved 27 November 2019

Yin, Junjun; Soliman, Aiman; Yin, Dandong; and Wang, Shaowen: *Depicting urban boundaries from a mobility network of spatial interactions: a case study of Great Britain with geo-located Twitter data*. International Journal of Geographical Information Science, 31(7), 2017, 1293–1313. doi: 10.1080/13658816.2017.1282615