

## ARTICLE

# Mapping cognitive place associations within the United Kingdom through online discussion on Reddit

Cillian Berragan<sup>1</sup>  | Alex Singleton<sup>1</sup>  | Alessia Calafiore<sup>2</sup>  | Jeremy Morley<sup>3</sup> 

<sup>1</sup>University of Liverpool, Liverpool, UK

<sup>2</sup>University of Edinburgh, Edinburgh, UK

<sup>3</sup>Ordnance Survey, Southampton, UK

## Correspondence

Cillian Berragan, University of Liverpool, Liverpool, UK.

Email: [c.berragan@liverpool.ac.uk](mailto:c.berragan@liverpool.ac.uk)

## Funding information

Economic and Social Research Council, Grant/Award Number: CDT (ES/P000401/1)

## Abstract

This paper explores cognitive place associations; conceptualised as a place-based mental model that derives subconscious links between geographic locations. Utilising a large corpus of online discussion data from the social media website Reddit, we experiment on the extraction of such geographic knowledge from unstructured text. First we construct a system to identify place names found in Reddit comments, disambiguating each to a set of coordinates where possible. Following this, we build a collective picture of cognitive place associations in the United Kingdom, linking locations that co-occur in user comments and evaluating the effect of distance on the strength of these associations. Exploring these geographies nationally, associations were shown to be typically weaker over greater distances. This distance decay is also highly regional, rural areas typically have greater levels of distance decay, particularly in Wales and Scotland. When comparing major cities across the UK, we observe distinct distance decay patterns, influenced primarily by proximity to other cities.

## KEYWORDS

cognitive place connectivity, distance decay, geoparsing, mental maps, natural language processing, United Kingdom

## 1 | INTRODUCTION

The importance of relational thinking to understand geographical phenomena has been widely acknowledged in human and computational geography (Bergmann & O'Sullivan, 2018; Glückler & Panitz, 2021; Lukermann, 1961). Spatial networks have been explored from a variety of perspectives to uncover the dynamics underpinning the spatial behaviours of individuals (González et al., 2008; Noulas et al., 2011), or to challenge conceptualisations of regions as bounded by administrative definitions (Alessandretti et al., 2020; Calafiore et al., 2021).

Within computational geography, most research has explored direct connections between places by investigating physical movements of individuals, using population movement data from both traditional data sources such as Census or surveys (Rae, 2009; Titheridge et al., 2009), or through alternative forms of data like transport records (Allard & Moura, 2016; Farber & Li, 2013; Gong et al., 2021; Yang et al., 2019), mobile phone data (Lin et al., 2019; Rowe, Lovelace, et al., 2022;

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

The information, practices and views in this article are those of the author(s) and do not necessarily reflect the opinion of the Royal Geographical Society (with IBG).

© 2024 The Authors. *Transactions of the Institute of British Geographers* published by John Wiley & Sons Ltd on behalf of Royal Geographical Society (with The Institute of British Geographers).

SafeGraph, 2022), and geotagged social media (Arthur & Williams, 2019; Li et al., 2021; Ostermann et al., 2015; Steiger et al., 2015). However, focusing only on connections built through population movement conceals associations that persist through individuals or community subconsciously, regardless of any physical movement.

Literature discussing the role of human cognition in constructing mental images of cities (Lynch, 1964), and how they can be represented through mental maps (Gould & White, 1986), reveals that the way humans conceive spatial structures and associations between places are substantially entrenched in individuals' experiences and geographic knowledge, which only partially derive from movements. Places represent a complex network of socio-spatial relationships that emerge from linked individual experiences (Pierce et al., 2011), enabling the definition of collectively recognised place associations. While movements in geographic space are limited by time and distance (Miller, 2018; Patterson & Farber, 2015), representational spaces expressed through mental maps are not necessarily bounded by spatio-temporal constraints (Merrifield, 1993). Modern developments in transport and communication access warp the perceptions of distance between places (Massey, 2008), and in turn their perceived level of connectivity (Fabrikant et al., 2002).

Alternatively, online sources of data offer novel opportunities to explore place associations, built directly from the passive contributions of individual users. Recent work has demonstrated how digital social friendships (Bailey et al., 2018), or embedded links in Wikipedia articles (Salvini & Fabrikant, 2016), may be used to provide insight into social place connections. Other works have instead considered that text itself can be used to quantify relationships between geographic terms, described as 'geo-semantic relatedness' (Ballatore et al., 2014). Work building on this concept has applied it to city and region names identified in news articles, social media and general web pages (Hu et al., 2017; Liu et al., 2014; Meijers & Peris, 2019; Ye et al., 2021).

Distance is a key influence on observed levels of connectivity in spatial interaction literature (Haynes & Fotheringham, 1985), and Tobler's first law of geography, where locations that are further apart are typically less well connected (Tobler, 1970), has generated the term 'distance decay' (Taylor, 1983), which has various forms of mathematical representation. Given a legacy of empirical evidence, distance decay in its various forms can be sensibly assumed in place connections when both temporal and spatial constraints are considered in our physical environment. However, when considering the links between locations from the perspective of cognitive associations built through mental maps, such constraints are no longer as restrictive (Fabrikant et al., 2002). Quantifying the effect of distance on cognitive place associations may therefore result in unexpected patterns in the effect of distance on associations that reveal the cognitive biases used to construct mental maps.

The objective of this paper is to quantify cognitive place associations across the UK<sup>1</sup> to build mental maps, while evaluating the effect of distance on the strength of these associations, measuring the level of distance decay through a gravity model. To generate association measures from a cognitive rather than geographic perspective, we infer associations through co-occurring locations extracted from a large corpus of informal, unstructured and discursive text from the social media website Reddit. Locations when mentioned in informal comments are drawn from a cognitive process associated with mental maps of these locations, subconsciously illustrating associations between places from memory and based on experience.

Section 2 outlines existing literature relating to cognitive place associations, detailing methods that can be used for the automated extraction and grounding of place names<sup>2</sup> from a large corpus of unstructured text. Section 3 provides details on our data sources, our methodology for geoparsing place names, and the computation of a gravity model to examine the effects of distance decay on the strength of associations. In Section 4 we present the results of our gravity model and demonstrate variations in distance decay with respect to six locations. In Section 5 we conclude our findings and outline the scope for future work.

## 2 | COGNITIVE PLACE ASSOCIATIONS

The term 'mental map' typically refers to the cognitive visualisation of a geographic environment. It represents collective, experiential geographic knowledge, relating to both places, and the relationships between them (Kaplan, 1976). Mental maps exhibit a variety of biases; for example, they are often more detailed with respect to locations that we are familiar with, while others may be less detailed or even absent entirely (Gould & White, 1986). The scale and distance between features in mental maps can be warped (Marston et al., 2005; Peake & Moore, 2004); prominent roads may appear larger than in reality, skyline features in a city may be perceived as less prominent due to their irrelevance to an individual at street level (Gould & White, 1986; Lynch, 1964), and good transport connections narrow the time it takes to reach connected locations, which in turn reduces the perceived distance between them (Massey, 2008;

Merrifield, 1993). Intermediate features along common routes also have varying levels of importance to individuals; unimportant features may appear less prominent than in reality or absent altogether (Carr & Schissler, 1969; Kaplan, 1976).

The characterisation of mental maps has been well studied from a qualitative perspective, often featuring individual participation for the physical construction of hand-drawn sketches (Canter, 1977; Goodey, 1974; Gould & White, 1986; Haney & Knowles, 1978; Lee, 1973; Lynch, 1964; Montello et al., 2003; Murray & Spencer, 1979; Pocock, 1976). Such approaches typically consider more localised areas that are familiar to selected participants (Canter, 1977; Haney & Knowles, 1978; Pocock, 1976), focusing on mapping landmarks and regions within cities. Others have considered the broad characterisation of larger regions like entire countries (Goodey, 1974; Gould & White, 1968), where mental maps are less detailed, instead contributing generalised information regarding areas that are deemed important to the participant. Inherently, these techniques capture subjective information from individuals, which may not necessarily conform with a generalised collective knowledge of these geographic locations.

## 2.1 | Quantifying associations

Within mobility research, connections between locations are broadly quantified through mapping the flow of populations, goods, services or other entities between origin and destination locations (Shaw & Hesse, 2010). This relates to the concept of *spatial interaction*, which describes a mathematical or statistical representation of physical movements over space, typically observed in the context of commuting, migration or information and commodity flows (Dennett & Wilson, 2013; Haynes & Fotheringham, 1985; Rowe, Calafiore, et al., 2022; Rowe, Lovelace, et al., 2022; Shaw & Hesse, 2010; Singleton et al., 2012). In spatial interaction literature, the strength of connectivity between locations is generally quantified through gravity models, which incorporate the effect of relative distance on the strength of interaction between origin and destination locations (Erlander, 1980; Haynes & Fotheringham, 1985). Typically, an increase in distance leads to a decrease in spatial interaction, known as 'distance decay' (Taylor, 1983). While conceptualising geographic connections in this manner builds a picture that is constrained by spatio-temporal movement, this is not necessarily a requirement for a more generalised understanding of associations as a geographic concept (Merriman, 2012).

Unlike physical connections, which are described by movements across Euclidean space, cognitive associations are decoupled from the restriction of physical movements; both the distance between locations, and the time taken to travel between them, do not directly influence the strength of cognitive associations. Instead, they capture the persistent perceived associations that reflect the experiential geographic knowledge used by individuals to generate 'mental maps' (Gould & White, 1986). Such associations capture non-specific subconscious links between locations, influenced by personal experiences, incorporating cultural similarities (Greenberg Raanan & Shoval, 2014), distortion through commuting methods and navigation technologies (Peake & Moore, 2004), online communication (Zook, 2006), and other influences on cognitive bias. For example, transport access and telecommunication warp a general sense of perceived geographic distance between certain locations (Massey, 2008), and the cognitive understanding of 'nearness' does not necessarily correlate with Euclidean distance (Fabrikant et al., 2002; Montello, 1993; Worboys, 2001). These implicit associations between places are generated through a complex network of socio-spatial relationships, built through linked individual experiences, and allow for shared experiences of places to be captured (Pierce et al., 2011).

Traditional approaches to the exploration of differences between cognitive and real-world associations would have relied on the use of large-scale studies and active individual participation to derive these associations, for example through volunteered geographic information (VGI; Goodchild, 2007) or participatory mapping (Chambers, 2006; Pánek, 2016). Additionally, while mental maps may be generated through hand-drawn sketches, such methods do not scale well to derive a population level understanding.

By contrast, alternative forms of data online present an opportunity to infer associations between places by capturing persistent links that are not temporally or spatially bounded. Facebook, for example, has been used to generate social connections between locations, geographically grounded based on the home location of two friends (Bailey et al., 2018). Geographic networks may also be generated from crowdsourced databases like Wikipedia (Salvini & Fabrikant, 2016), which demonstrate connections between cities based on hyperlinks embedded in articles that contain general knowledge. Unlike these structured data sources, unstructured online text also provides embedded geographic information as place names, which may be extracted through computational techniques, such as natural language processing (Berragan et al., 2022; Purves et al., 2018).

Relationships between locational mentions in text are typically examined through co-occurrences, where locations that are frequently mentioned in a shared context are assumed to have a real-world relationship (Ballatore et al., 2014; Hu et al., 2017; Liu et al., 2014; Meijers & Peris, 2019; Ye et al., 2021). However, current research has concentrated primarily on examining the relationships between city names on news articles (Hu et al., 2017) or general web pages (Liu et al., 2014), where locational mentions do not necessarily capture a collective and generalised view generated from the mental perceptions of geography that exist within populations.

Alternative sources include online social media, which contribute a large volume of natural language text submitted by many unique users, discussing a range of informal topics, typically with shared user interactions. Place names discussed on social media more frequently include fine-grained locations (Han et al., 2018; Li & Sun, 2014), and given interactions are often more informal, the information captured likely exhibits user cognitive biases related to their own mental maps (Jang & Kim, 2019). While past work has built co-occurring place names from single news articles or documents, they may instead be built from a user-facing perspective, building co-occurrences from comments associated with each user in a large corpus of social media data. We argue that this approach more appropriately captures the cognitive information each user utilises to associate two locations, which can then be generalised by combining associations across each user in the corpus.

However, there are concerns with the use of passively contributed, user-generated data for place-focused geographic research; primarily with the representativeness of populations and the bias in contributions (Gardner et al., 2020; Graham et al., 2015). For example, despite having over 300 million users, Twitter users typically post from high-density urban areas rather than where they live (Ballatore & De Sabbata, 2020), demographic groups have variable propensity to contribute (Ballatore & De Sabbata, 2018; Gardner et al., 2020; Hecht & Stephens, 2014), and contributions to gazetteers or digital maps are increased in more densely populated, urban locations (Graham et al., 2015; Laurier et al., 2016; Smith et al., 2020). Another concern with user-generated data comes from the tendency for few users to contribute the greatest proportion of activity (Haklay, 2016), meaning that despite a large volume of unique users, there may be bias towards the contributions of certain individuals. In other research methods like participatory mapping, biases often reflect the social and cultural background of the communities contributing their understanding of geographies (Corbett & Rambaldi, 2009; Pánek, 2016), which in our work equates to the experiential knowledge used to construct those mental maps that inform our cognitive place associations.

## 2.2 | Extracting locations from text

Past works that considered links between locational mentions in text have identified locations either by querying articles for city names (Hu et al., 2017) or simply using a word list of city names to parse articles for their occurrences (Meijers & Peris, 2019). Such approaches suffer with performance; Meijers and Peris (2019) for example identified that 2.8% of their target place names could refer to multiple locations, while 1% of names were words that appeared in the English vocabulary. In total, they identify that around 15% of their place names displayed some level of ambiguity, and quantitative assessment of the effect of this demonstrated that it negatively impacted the quality of the associations identified. To avoid such issues, instead of a simple rule-based approach for the extraction of place names from our text, we construct a structured process using machine learning. This implements geoparsing, which is the process of extracting place names from unstructured text and matching them to the correct associated geographic coordinates (Purves et al., 2018). This task can be divided into two stages; identifying place names in text, followed by the association of these place names with a unique identifier in a knowledge base (typically a gazetteer) in a process called geocoding or toponym disambiguation.

Modern geoparsing processes use Named Entity Recognition (NER) to identify place names from natural language text (Halterman, 2017; Karimzadeh et al., 2019; Purves et al., 2018). Unlike simpler methods which use knowledge or rule-based methods (Leidner & Lieberman, 2011), NER uses more complex supervised machine learning to identify place names. The use of machine learning allows for the identification of place names that do not already appear within formal gazetteers, which is particularly useful in research considering colloquial names (Hollenstein, 2008). Word context may also be used to improve accuracy, as words may appear in a gazetteer but not be used in a geographic context (Reading could be considered a place in the UK or a noun; Purves et al., 2018). This is particularly important when considering informal text, where capitalisation may not always indicate the use of proper nouns, misspellings may be frequent, and names that do not often appear in gazetteers are common.

However, recent work has noted that current geoparsing systems using existing NER models do not necessarily perform well for the task of place name extraction (Berragan et al., 2022). Such pre-built models do not always consider



geographically specific issues like the use of metonyms (Gritta et al., 2020), and are typically trained on news articles, which limits their performance on other forms of text, like social media (Berragan et al., 2022; Won et al., 2018). For toponym disambiguation, the global GeoNames<sup>3</sup> database is typically used as a gazetteer in these geoparsing systems, which has limited data for fine-grained locations in the UK (Moncla et al., 2014; Stock et al., 2013), while increasing potential noise with the inclusion of place names outside the UK. As such, existing geoparsers were considered unsuitable for our task; geoparsing UK place names within Reddit comments, with the inclusion of fine-grained locations.

### 3 | METHODOLOGY

We first developed a task-specific geoparsing process to identify all place names contained within the Reddit comment corpus, resolving them to geographic coordinates within the UK. Cognitive associations were then generated between each identified location using co-occurrence: when identified locations are mentioned by the same author, they create an association between places. These associations are therefore built from the mental maps of individual Reddit users, unbounded from the typical space–time constraints of traditional spatial relationships derived from movements. We then investigate the strength of these associations by deriving aggregate geographic representations derived from each user, and determine the role of distance in shaping the strength of these associations through a gravity model. Finally, we select four urban and four rural regions to map the strength of associations geographically, demonstrating variations in distance decay patterns.

#### 3.1 | Geoparsing Reddit comments

Reddit<sup>4</sup> is a public discussion, news aggregation social network, and among the top 20 most visited websites in the UK. As of 2020, Reddit had around 430 million active monthly users, comparable to the number of Twitter users (Murphy, 2019; Statista, 2022). Reddit is divided into separate independent *subreddits* each covering specific topics of discussion, where *users* may submit *posts* that have dedicated nested conversational threads enabling users to add and respond to *comments*. Subreddits cover a wide range of topics, and in the interest of geography, they also act as forums for the discussion of local places. The UK subreddit<sup>5</sup> acts as a general hub for related topics, notably including a list of smaller and more geographically specific related subreddits. This list provides a ‘Places’ section, a collection of local British subreddits, ranging in scale from country level (/r/England), regional (/r/thenorth, /r/Teeside), to cities (/r/Manchester) and small towns (/r/Alnwick). In total there are 213 subreddits that relate to ‘places’ within the UK.<sup>6</sup> For each subreddit, every single historic comment was retrieved using the Pushshift<sup>7</sup> Reddit archive (Baumgartner et al., 2020). In total 8,070,827 comments were extracted, submitted by 490,534 unique users, between 1 January 2011 and 17 April 2022, which represents a very large corpus of text comprising 262 million words.

We then implemented our own geoparsing methodology to extract and geolocate any place name mention within each comment text. We first identified all place name mentions using a custom-built NER model.<sup>8</sup> This model was built using a large language model called BERT (Devlin et al., 2019), which is pre-trained on a large corpus of general human text, meaning for tasks like NER it performs better compared with simpler models. Our NER model was then trained to identify all place names within this corpus. Coordinate information was attributed with all identified place names using OS Open Names,<sup>9</sup> and ‘natural’ locations from the Gazetteer of British Place Names.<sup>10</sup> Given place names typically appear multiple times in gazetteers, a disambiguation method was required. We disambiguated place names by finding their minimum distance to a collection of contextual locations. Contextual locations in this case referred to all gazetteer entries matching place names that appear in sentences with this target place name, within the same subreddit. This worked under the assumption that each unique place name in a single subreddit is likely to refer to the same location, and that locations mentioned in surrounding text are likely geographically close together (Kamalloo & Rafiei, 2018). When associating locations with coordinate information, we excluded any location that was larger than a city, for example countries or regions.

Our final dataset consisted of a collection of place names with their geographic coordinates, corpus location, and an anonymised user ID for the user of the comment the place name was taken from. In total, 213,764 unique users mentioned at least one place name in our corpus, 39,050 mentioned more than 10 place names, and 3158 over 100. One percent of these contributed 32% of all place names, representing the top 2137 users. As is common in user-generated content, our data are skewed in that proportionally few users mention a large proportion of our total place names.

However, the large volume of unique users that contribute low volumes of comments does mean that we likely still achieve a broad representation, particularly compared with past work that generated mental maps for a limited number of individuals (Goodchild & Li, 2012). As our comments spanned a period of over 10 years, we also examined the temporality of contributions made by users. The mean time between a user's first and final comment is 318.1 days, with a maximum of 4112 days. As such, the contributor distribution is highly skewed as the majority of users (55%) only have commented a maximum of 1 day apart.

### 3.2 | Place associations through co-occurrence

'Cognitive association strength' is defined in our paper as the normalised proportion of co-occurrences between two locations in our corpus, where co-occurrences represent the total collection of locations mentioned by a single user. The following section first outlines the construction of distance decay measures using a gravity model that incorporates cognitive association strength alongside distance, then details how we generate a scaled measure of this cognitive association strength. The first measure enables us to quantify how distance impacts our association strength, determining whether there is an observable distance decay effect when considering locational co-occurrences in user comments. The second enables the direct strength of association between locations to be examined, without the incorporation of distance in the calculation.

To measure the effect of distance decay we employ the same gravity model used by both Liu et al. (2014) and Hu et al. (2017), shown in Equation (1):

$$S_{ij} \propto \frac{S_i S_j}{d_{ij}^\beta} \quad (1)$$

where  $S_{ij}$  is the total number of users that mention both places  $i$  and  $j$ , and  $S_i S_j$  is the total number of users that mention place  $i$  multiplied by the total number of users that mention place  $j$ .  $d_{ij}$  is the distance between the two locations  $i$  and  $j$ , and  $\beta$  is the friction factor. Larger values for  $\beta$  indicate a stronger distance decay effect. Estimating the value of  $\beta$  generates a quantifiable measure of the distance decay effect (Hu et al., 2017).

We can decompose Equation (1) into the following multiple linear regression model (Taaffe, 1996):

$$\log(S_{ij}) = b_0 + b_1 \log(S_i S_j) + b_2 \log(d_{ij}) \quad (2)$$

where  $b_2 = -(\beta * b_1)$ , meaning we can calculate our  $\beta$  coefficient using  $\beta = -(b_2/b_1)$  (Hu et al., 2017; Li & Sun, 2014).<sup>11</sup>

While this approach enables the calculation of a global  $\beta$  to measure distance decay, a spatial regression model would enable us to calculate local values of  $\beta$ , quantifying the distance decay effect on individual locations (Rey et al., 2023). Therefore, we additionally implement a spatial regression model which incorporates a fixed spatial effect for the H3 polygon name, allowing for  $\beta$  coefficients to be calculated for each location in our study, to explore spatial heterogeneity.

Finally, we generate a normalised cognitive association measure to assess the strength between two locations. Unlike the previous gravity models, co-occurrences are not incorporated alongside distance. This mirrors similar work that considered the strength of social connections between Facebook (Bailey et al., 2018) and Twitter users (Li et al., 2021), enabling the direct strength of associations to be generated:

$$\frac{S_{ij}}{\sqrt{S_i S_j}} \quad (3)$$

In this equation, dividing by  $\sqrt{S_i S_j}$  normalises our values, given locations with higher populations are expected to be mentioned by a larger number of users. Therefore, values range from 0, indicating no association, to 1, showing a complete overlap in user mentions.

To present the results of our analysis we aggregate our user location mentions into H3 hexagons,<sup>12</sup> a hierarchical spatial indexing system which partitions all locations across earth into a uniform hexagonal grid, available for different levels of aggregation. We select an H3 resolution of 5, which equates to an average hexagon area of 252 km<sup>2</sup> and an edge length of 9.8 km. All associations between each location contained within a shared H3 hexagon are then combined,

forming association measures between hexagons, rather than unique point locations. To name hexagons we select the most frequently occurring location.

The use of fixed unit size hexagons for aggregating data in our analysis is beneficial for several reasons. Firstly, hexagons are geometries that enable us to obtain results that are statistically more robust, especially when analysing distance decay between locations, because of the constant number of neighbours, with an equal distance separating them (Birch et al., 2007). Secondly, hexagon grids help to minimise misrepresentation in spatial visualisation (Langton & Solymosi, 2021), and allow us to capture inter-region heterogeneity. Finally, aggregation is essential given the data representations of locations within gazetteers; despite many locations having large footprints, they are all represented as a single coordinate pair (Goodchild & Hill, 2008). This problem means that despite users mentioning locations like parks within cities, without aggregation they are treated as two distinct points, with a geographic distance separating them. Alternatively OpenStreetMap can be used to provide more accurate place footprints, but at the cost of a very large data volume when considering the entirety of the UK (Haklay & Weber, 2008).

Additionally, we classify our H3 hexagons into both rural and urban using the England and Wales Rural Urban Classification<sup>13</sup> and Scotland Rural Urban Classification.<sup>14</sup> For Scotland, classes 1 and 2 were considered urban.

## 4 | RESULTS

In the following section, we first examine the performance of our geoparsing methodology, identifying any potential noise and how this was mitigated. We then examine our cognitive place associations, exploring how distance impacts the strength of association by generating gravity models to calculate  $\beta$  coefficients both globally and locally. Finally, we examine the patterns in association strength across a selection of targeted geographic locations.

### 4.1 | Extracting names and locations: Assessing geoparsing performance

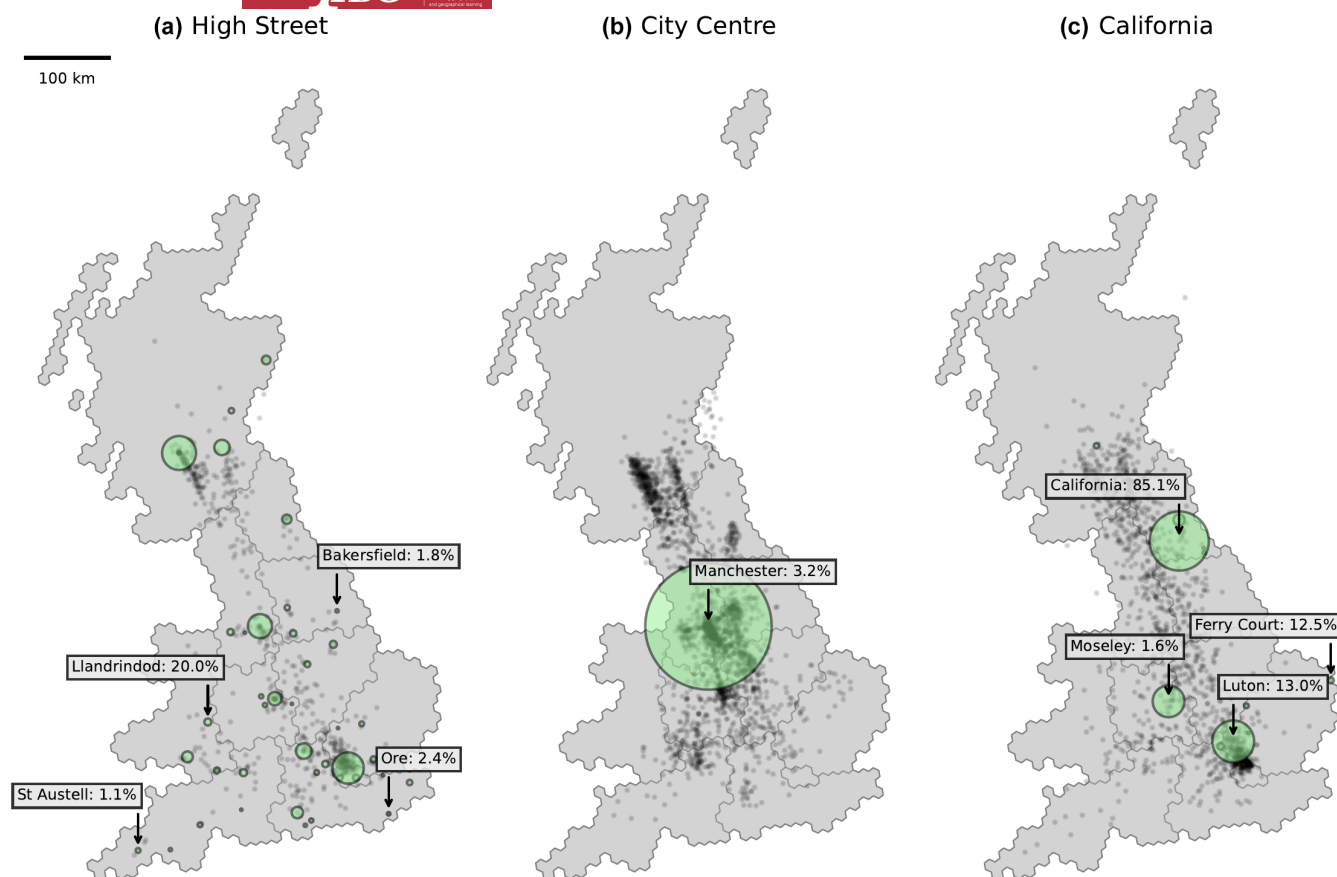
In total, 26.8% of all comments within the Reddit corpus contained at least one place name: 5,001,261 place names were identified, with 2,848,310 (57.0%) being attributable to a set of coordinates.<sup>15</sup> From these locations, 42,333 were found to be unique, of which 21,014 were only mentioned a single time, while London was the most frequently encountered location, at 283,521 mentions. The most ambiguous place name was found to be 'High Street', with 47 total unique coordinate locations. As expected, many of the most ambiguous place names were street names, including 'Church Street' (36 locations), 'London Road' (34 locations) and 'Bridge Street' (34 locations).

In Figure 1 we consider three examples where place names may have been incorrectly geoparsed. Figure 1a shows the geographic distribution of all 47 'High Street' locations. The percentage values indicate the proportion of 'High Street' mentions within a particular H3 polygon compared with all other locations in this polygon. Aggregation here appears to mitigate the risk of noise in most cases, given most 'High Street' locations contribute less than 1% towards polygon associations. A similar case is shown in Figure 1b, where 'City Centre' mentions only account for 3.2% of the Manchester hexagon. Figure 1c instead demonstrates a location that is impossible to correctly geoparse in our model, and despite there being 13 unique locations in the UK called 'California', this issue only appears prominent in one hexagon. This hexagon named 'California' does potentially generate noise in our analysis, given the high contribution of 85.1%. However, as users that mention 'California' are spread across the country, it is less likely to largely impact our associations.

Notably, despite both 'High Street' and 'City Centre' being shared with non-specific geographic concepts, the model is still able to distinguish between them depending on context. For example, 'city centre' appears 23,961 times in our corpus, but is only tagged by our NER model 3008 times. While 'high street' appears 7773 times and is only tagged 768 times. These results suggest that the model is often able to correctly understand that identical phrases may or may not refer to place names, depending on their semantic context.

### 4.2 | Measuring distance decay of cognitive association strength

In the following section we present the levels of distance decay observed when evaluating place association strength through the gravity model specified in Equation (1), quantifying the level of distance decay using a  $\beta$  coefficient. As calculated, higher  $\beta$  coefficient values indicate a stronger distance decay effect, meaning that co-occurrences between



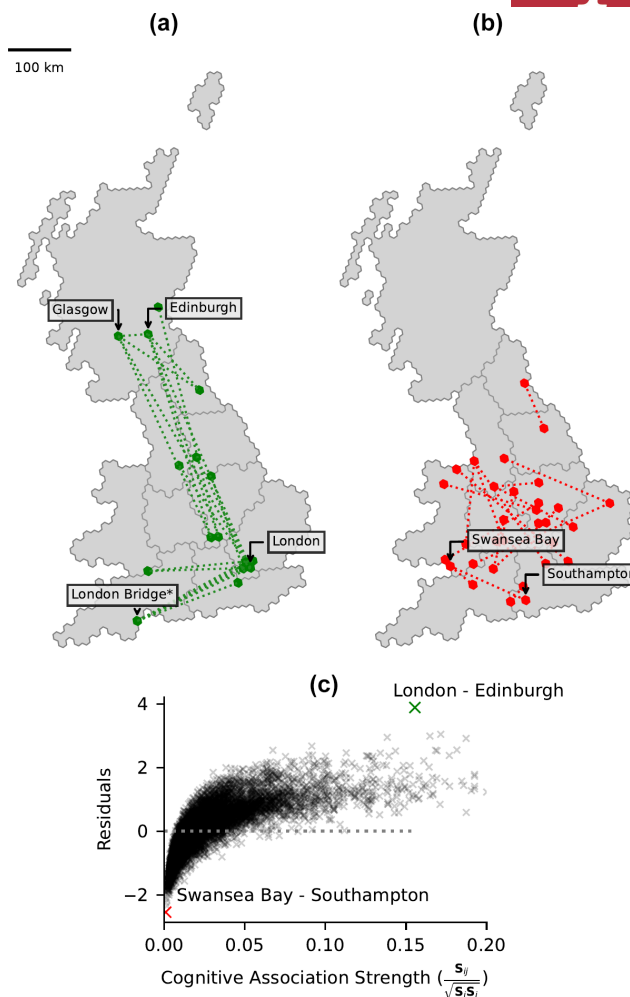
**FIGURE 1** Locations of three place names that appear in the UK gazetteer that are difficult to correctly disambiguate; (a) High Street, (b) City Centre and (c) California. Size of the green points indicates frequency in mentions, black points are user locations determined through mean locational mentions. Values indicate the proportional contributions of each disambiguated location to their respective polygon (top four percentages shown).

locations that are geographically more distant tend to be less frequent. A  $\beta$  value of zero would indicate that distance has no effect on the frequency of co-occurrences between locations.

Our gravity model gives a  $\beta$  coefficient of 1.00 (Pearson's  $R^2$  0.772), reflecting a distance decay from co-occurrences in Reddit comments that is stronger than decay observed in other studies that explored news articles (0.23; Hu et al., 2017), or general web queries (0.2; Liu et al., 2014). Confirming the existence of a general distance decay effect for Reddit derived places demonstrates that distance typically contributes to lower co-occurrences in locations that are further apart, a similarity that is shared with past work that examines decay from the perspective of true population movements (Gong et al., 2021; Yang et al., 2019), and the social relationships of regions examined through social media (Bailey et al., 2018; Li et al., 2021). Our place associations generated from users on Reddit therefore appear to more appropriately incorporate a geographic component compared with city mentions in news articles or general web pages. However, while this gravity model gives us an indication of the global level of distance decay in our corpus, it is likely that the level of distance decay varies by location. In the following analysis we consider locations where the gravity model does not achieve a good approximation.

Figure 2a,b plots the top (most positive) and bottom (most negative) 20 residuals from our gravity model, demonstrating associations that are stronger or weaker than expected when accounting for the distance between two regions. Many of the top residuals concern associations shared with London and other major cities in the UK, with some associations between urban areas in Scotland. The most positive residual is the association between London and Edinburgh (3.89), Glasgow and London in second (3.84), and Glasgow and Edinburgh in third (3.81). As expected, these residuals reflect a strong association between regions over larger distances (mean 293 km), highlighting associations where distance decay is less effective. Notably, there is an incorrect association here between a natural feature named 'London Bridge' and London (Figure 2a, see London Bridge\*), which has appeared due to the lack of urban landmarks in our gazetteer. The bottom residuals are more sporadic, typically showing associations over shorter distances (mean





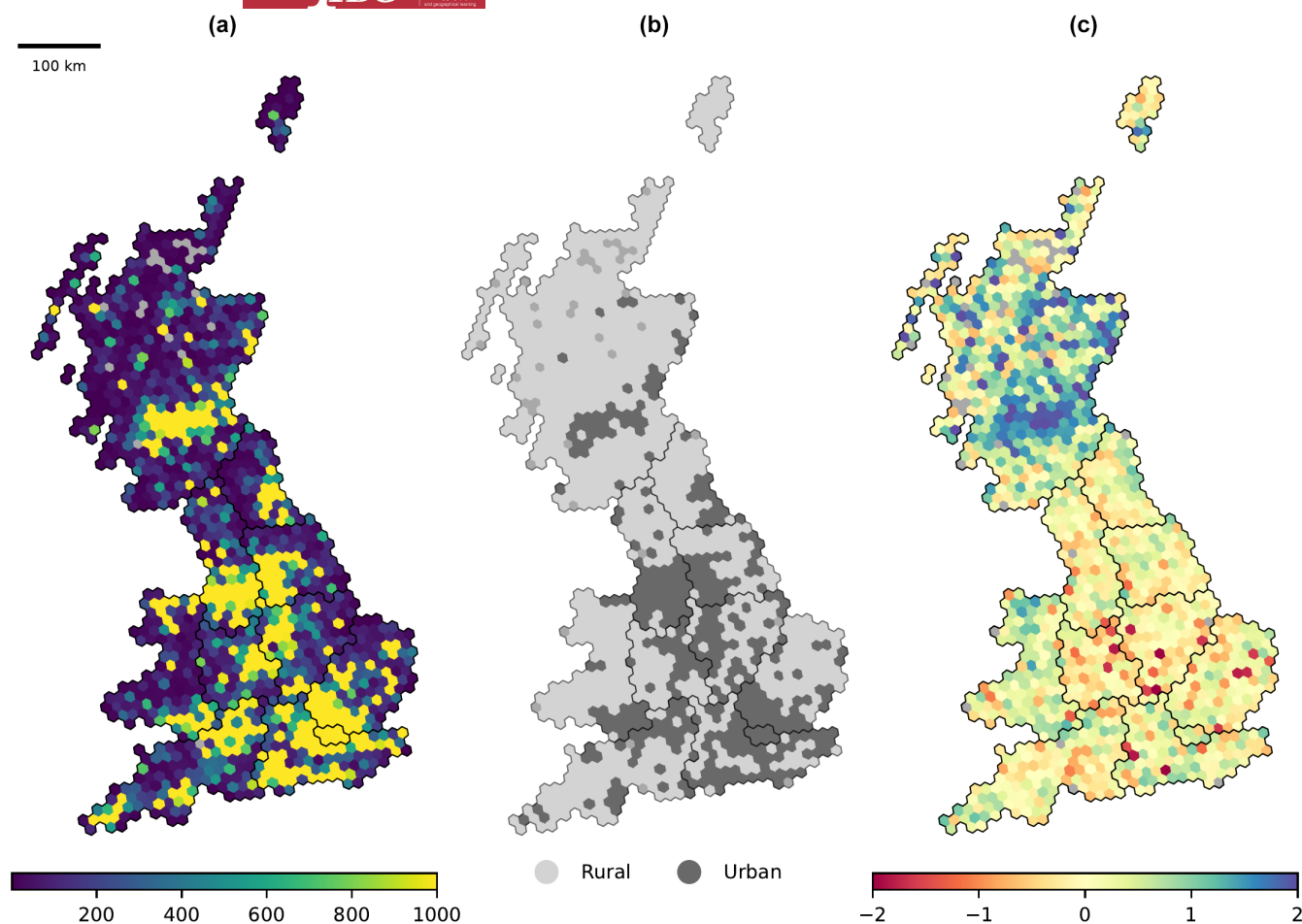
**FIGURE 2** H3 polygons showing (a) top 20 associations by residual values in green ( $>0$ ), and (b) bottom 20 associations by residual values ( $<0$ ) in red. (c) Residuals taken from Equation (2) against co-occurrence strength (10,000 samples). (\*) Indicates a location that is incorrectly geoparsed.

156 km) between lesser known locations that are unusually weak. For example, highlighted in this figure is the association between Swansea Bay and Southampton ( $-2.55$ ). Figure 2c plots the model residuals against cognitive association strength, showing that for locations with a greater proportion of co-occurrences, the model is likely to be under-estimating in prediction, leading to an over-prediction in distance decay, with the inverse true for locations with a lower proportion of co-occurrences.

### 4.3 | Regional difference in distance decay

Figure 2 demonstrates that there are clear regional variations in the observed level of distance decay, which do not conform with the general decay effect calculated through our proposed gravity model approximation. To examine regional effects on distance decay, we implement a spatial regression model (a mixed linear model with spatial fixed effects), allowing  $\beta$  values to change depending on the location of each polygon.

Incorporating this spatial information gives a more effective approximation of our gravity model, and achieves an improved Pearson's  $R^2$  of 0.946, suggesting that distance decay is not uniform across all regions in our study. Figure 3a shows the distribution of locations mentioned in our study, which broadly conform with the binary urban rural classification shown in Figure 3b. Figure 3c maps the spatial  $\beta$  coefficients obtained through our spatial regression model, with high distance decay present across Scotland, Wales, and areas in the South West and North East of England. Users that mention locations in these regions typically do not mention other locations that are geographically distant, highlighting areas that are either more isolated from the rest of the UK, or have stronger associations with nearby locations.



**FIGURE 3** H3 polygons showing (a) distribution of geoparsed locations. (b) Urban rural classification index for England, Wales and Scotland, reclassified into binary ‘Urban’ or ‘Rural’. (c) Calculated  $\beta$  coefficients for the spatial regression model; higher  $\beta$  values indicate a greater distance decay strength.

In the North East, this perceived isolation from the rest of the UK mirrors lexical research, where Tweets in the North East have been shown to be unlike other regions (Arthur & Williams, 2019). This region in particular has been known to suffer economically following the historic decline of local industries (Middleton & Freestone, 2008), where lack of job opportunities has resulted in poor inward migration, with among the lowest population growth in the country (Office for National Statistics, 2022). Alternatively, this observation may also be attributable to a general sense of identity that is associated with these regions. Both the South West and North East of England are known to exhibit a strong sense of localised identity (Deacon, 2007; Middleton & Freestone, 2008), which is similarly translatable to the national identity that generates strong associations within Scotland and Wales that are not shared with England (Haesly, 2005).

To explicitly quantify the difference in distance decay between urban and rural areas we calculate separate  $\beta$  coefficient values based on the binary split of areas into urban or rural. Urban areas have a  $\beta$  coefficient of 0.69, while rural areas have a  $\beta$  coefficient of 1.13, indicating that urban areas do appear to have a lower overall level of distance decay compared with rural regions. This correlates with the results of traditional mobility studies where more populated areas tend to exhibit a lower distance decay (Thomas, 1981), largely dictated by the improved accessibility to external locations through public transport, the road network or job opportunities (Findlay et al., 2000; Moseley, 2023), and the general cultural significance that is more frequently associated with urban locations (Borer, 2006; Lynch, 1964).

We have demonstrated that not only does distance decay vary between rural and urban locations, but within these classes there is also apparent heterogeneity. In the following section, we consider the ability to directly map the strength of cognitive associations with respect to a selection of both rural and urban regions in our study, to understand the variation in distance decay patterns.

## 4.4 | Mapping cognitive place associations

In Figure 4 we map the cognitive association strength of each H3 polygon in our study with respect to four major cities and four rural locations in the UK, also indicating the associated  $\beta$  coefficients. Mapped cognitive association values are given by Equation (3), and indicate the proportion of users that have comments that mention locations both within the target polygon (e.g., London), and locations in other polygons. Distance decay curves for each polygon are shown below these maps, indicating patterns in decay associated with each location. London has the lowest  $\beta$  coefficient of all cities, indicating that locations at increasing distance from London decay in their association at a slower rate compared with other cities. This is reflected by the shallow overall decay curve for London, increasing at points associated with main urban conurbations in England and Scotland, observable on the map for Figure 4a. Such trends are perhaps unsurprising given London's prominence as the capital city. Manchester in Figure 4b reveals a different decay pattern, showing a sharp drop in associations initially, which reduces and reverses when cities like London or Edinburgh are included in the distribution. Unlike Manchester, Newcastle (Figure 4d) has an overall greater  $\beta$  coefficient, where association strength drops more quickly and is less persistent across England, only increasing with urban locations in Scotland, and slightly with London. While both are major cities in England, Manchester is both physically more well connected to the rest of the country through existing rail routes (Miyoshi & Givoni, 2013), and is a greater economic centre compared with Newcastle. These factors likely contribute to the perceived strength of associations with these cities, which is captured in our analysis.

Edinburgh (Figure 4c) is distinct compared with other cities, with a steeper initial decay curve compared with London, largely dictated by stronger initial associations with Scottish locations. Again, as is common for many cities in the UK, this city also shares a strong association with London, regardless of the distance. This increased strength of association with locations within Scotland gives Edinburgh the highest  $\beta$  coefficient, an effect that captures the strong sense of identity between areas in Scotland (Haesly, 2005).

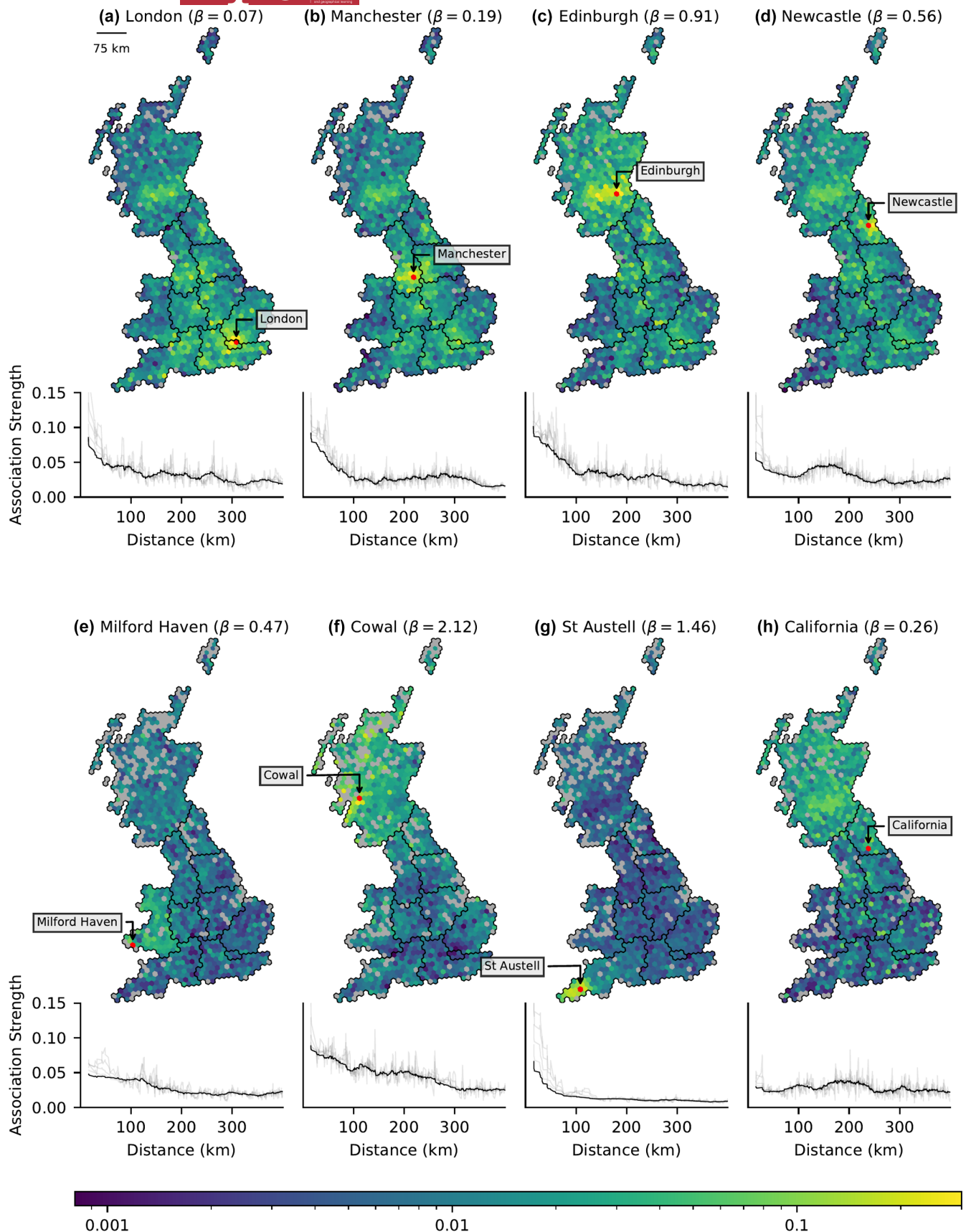
Figure 4e–g give examples of variable distance decay curves for rural locations across the UK. Both '(e) Milford Haven' in Wales, and '(f) Cowel' in Scotland share general associations across each respective country, which appears to drop past the border into England. This similarly captures the sense of national identity associated with both Wales and Scotland, and conforms with results from the analysis of both physical and online social networks, where strong 'boundary effects' often see intra-connectivity within regions that becomes weaker when moving across borders (Arthur & Williams, 2019; Bailey et al., 2018; Li et al., 2021; Yin et al., 2017). Given a national identity is less prominent in England, the town '(g) St Austell' gives a steep distance decay curve, with low association strength between any location more than 50 km away, a noticeably different curve compared with the rural locations analysed in Scotland and Wales.

We also examine the incorrectly disambiguated '(h) California' polygon, and confirm that the distance decay curve does not appear to show a geographically cohesive pattern, with no noticeable gradient. The positive  $\beta$  coefficient appears to relate to an unexplained increase in association with Scotland, however values remain low.

## 5 | CONCLUSIONS, IMPLICATIONS AND FUTURE WORK

In our work we present an alternative method for determining subconscious associations between locations, generating quantifiable measures of association strength solely using user-generated social media text. Unlike physical or online social interactions, our cognitive associations are intended as a persistent measure of strength between locations across the UK, built from the naive, place-based geographic knowledge of individuals. Our geoparsing process means that no explicitly geographic information like geotags are required in our data source, allowing for the inclusion of fine-grained and informally defined locations, and associations may be examined between any two locations identified in our corpus.

By utilising a distance-based gravity model, we demonstrate that our associations do broadly conform with established real-world geographic restrictions through an observed distance-decay effect, but with notable deviations. Unlike past work that only considered co-occurring city names, we expand our analysis to incorporate place name mentions of any scale, which enables the exploration of both rural and suburban decay. Therefore, we are able to demonstrate that distance decay is greater in rural areas compared with urban areas, and that cities across the UK have varying patterns in distance decay. Associations between major cities like Manchester and London are demonstrably stronger than less prominent intermediate locations, an effect that challenges the notion of Euclidean distance in a mental understanding of geography (Carr & Schissler, 1969; Kaplan, 1976), and conforms with the suggestion that 'nearness' is not a uniform concept (Fabrikant et al., 2002; Massey, 2008; Montello, 1993; Worboys, 2001). Alternative



**FIGURE 4** Data subsets with respect to four urban locations; (a) London, (b) Manchester, (c) Edinburgh and (d) Newcastle, and four rural locations; (e) Milford Haven, (f) Cowal, (g) St Austell, and (h) California, showing cognitive association strength associated with each H3 polygon containing each named location (highlighted in green),  $\beta$  values generated from data subsets. Distance decay plots below maps show association strength against distance for each selected location. Lines show rolling mean for 250 samples in black and lower samples in grey.



patterns are also demonstrated, particularly for locations in Scotland, where distance appears to have little impact on association strength until locations across the border into England are reached. This observation appears to correlate with the concept of a 'boundary effect' that has been captured in both physical and online networks (Arthur & Williams, 2019; Bailey et al., 2018; Li et al., 2021; Yin et al., 2017). This example is particularly interesting, replicating past research that generated mental maps from individuals through participatory mapping, which captured strong desirability towards areas within Scotland from residents that did not persist across the border into England (Gould & White, 1968).

The distinct patterns in distance decay that we observe demonstrate differences in associations between cities that capture real-world perceptions of these locations. For example, regions of low association across the UK may reflect the lack of desire to connect more broadly with the rest of the country (Roos Breines et al., 2019). This is particularly noticeable in Scotland, Wales and the North East of England, regions where populations often exhibit a strong sense of independence from the UK, driven by a sense of regional or national pride (Haesly, 2005; Middleton & Freestone, 2008; Nayak, 2016). Strong associations with London and Manchester are also indicative of their importance nationally, while cities like Newcastle are perceived as less important, resulting in lower associations. Imbalance in rural locations may also be driven by an increasing dependence on digital maps (Farman, 2020); rural landmarks are far less common compared with cities with named streets and buildings (Laurier et al., 2016), limiting our ability to conceptualise rural environments (Smith et al., 2020).

Finally, Reddit is unique compared with alternative social media sources; general activity is centred around discussing specific topics or themes within communities, relative to more general social networks such as Twitter or Facebook (Medvedev et al., 2019; Sylla et al., 2022). Communities on Reddit therefore present the opportunity to generate collective, but geographically disaggregated representations of spatial knowledge. Locations identified from within these communities likely represent urban areas of interest which may be derived based on their frequency of mentions (Chen, 2019), or semantic regions that reflect mental perceptions of places (Gao et al., 2017). These unstructured comments also provide contextual lexicons relating to places names, meaning there is also the opportunity to explore associations between these communities through their associated typology (Arthur & Williams, 2019; Gao et al., 2017). While count-based lexical approaches have been traditionally used to explore geographic variation in text, large language models are now able to capture deep contextual semantic information (Devlin et al., 2019), allowing for a deeper connection between language and geography to be explored.

## ACKNOWLEDGEMENTS

We would like to thank all the reviewers for their constructive comments that supported the publication of this paper. Thank you also to Ordnance Survey and the Economic and Social Research Council (ES/P000401/1) for financially supporting this work.

## DATA AVAILABILITY STATEMENT

The data and code that support the findings of this study are openly available in FigShare at <https://figshare.com/s/1ca488dbcc97a5fd3294>.

## ORCID

Cillian Berragan  <https://orcid.org/0000-0003-2198-2245>

Alex Singleton  <https://orcid.org/0000-0002-2338-2334>

Alessia Calafiore  <https://orcid.org/0000-0002-5953-2891>

Jeremy Morley  <https://orcid.org/0000-0002-3658-8796>

## ENDNOTES

<sup>1</sup> Ordnance Survey UK does not include data for Northern Ireland.

<sup>2</sup> In our work we consider 'place names' to be ambiguous noun phrases found in text, without associated coordinates, while 'locations' are place names that have attributed coordinates. We use 'place associations' to capture the vagueness involved in cognitive associations.

<sup>3</sup> <https://www.geonames.org>.

<sup>4</sup> <https://reddit.com>.

<sup>5</sup> <https://reddit.com/r/unitedkingdom>.

<sup>6</sup> [https://www.reddit.com/r/unitedkingdom/wiki/british\\_subreddits](https://www.reddit.com/r/unitedkingdom/wiki/british_subreddits).

- <sup>7</sup><https://pushshift.io/>.
- <sup>8</sup>[https://huggingface.co/cjber/reddit-ner-place\\_names](https://huggingface.co/cjber/reddit-ner-place_names).
- <sup>9</sup><https://www.ordnancesurvey.co.uk/business-government/products/open-map-names>.
- <sup>10</sup><https://gazetteer.org.uk>.
- <sup>11</sup>For a more detailed mathematical explanation, see Hu et al. (2017).
- <sup>12</sup><https://www.uber.com/en-GB/blog/h3/>.
- <sup>13</sup><https://www.gov.uk/government/collections/rural-urban-classification>.
- <sup>14</sup><https://www.gov.scot/publications/scottish-government-urban-rural-classification-2020>.
- <sup>15</sup>Note that many names absent from our gazetteer include locations outside the UK.

## REFERENCES

- Alessandretti, L., Aslak, U. & Lehmann, S. (2020) The scales of human mobility. *Nature*, 587(7834), 402–407. Available from: <https://doi.org/10.1038/s41586-020-2909-1>
- Allard, R.F. & Moura, F. (2016) The incorporation of passenger connectivity and intermodal considerations in intercity transport planning. *Transport Reviews*, 36(2), 251–277. Available from: <https://doi.org/10.1080/01441647.2015.1059379>
- Arthur, R. & Williams, H.T.P. (2019) The human geography of twitter: Quantifying regional identity and inter-region communication in England and Wales. *PLoS One*, 14(4), e0214466. Available from: <https://doi.org/10.1371/journal.pone.0214466>
- Bailey, M., Cao, R., Kuchler, T., Stroebel, J. & Wong, A. (2018) Social connectedness: Measurement, determinants, and effects. *Journal of Economic Perspectives*, 32(3), 259–280. Available from: <https://doi.org/10.1257/jep.32.3.259>
- Ballatore, A., Bertolotto, M. & Wilson, D.C. (2014) An evaluative baseline for geo-semantic relatedness and similarity. *GeoInformatica*, 18(4), 747–767. Available from: <https://doi.org/10.1007/s10707-013-0197-8>
- Ballatore, A. & De Sabbata, S. (2018) Charting the geographies of crowdsourced information in greater London. In: Mansourian, A., Pilesjö, P., Harrie, L. & van Lammeren, R. (Eds.) *Geospatial technologies for all*. Cham, Switzerland: Springer International Publishing, pp. 149–168. Available from: [https://doi.org/10.1007/978-3-319-78208-9\\_8](https://doi.org/10.1007/978-3-319-78208-9_8)
- Ballatore, A. & De Sabbata, S. (2020) Los Angeles as a digital place: The geographies of user-generated content. *Transactions in GIS*, 24(4), 880–902. Available from: <https://doi.org/10.1111/tgis.12600>
- Baumgartner, J., Zannettou, S., Keegan, B., Squire, M. & Blackburn, J. (2020) *The Pushshift reddit dataset*. arXiv. Available from: <https://doi.org/10.48550/arXiv.2001.08435>
- Bergmann, L. & O'Sullivan, D. (2018) Reimagining GIScience for relational spaces: Reimagining GIScience. *The Canadian Geographer/Le Géographe Canadien*, 62(1), 7–14. Available from: <https://doi.org/10.1111/cag.12405>
- Berragan, C., Singleton, A., Calafiore, A. & Morley, J. (2022) Transformer based named entity recognition for place name extraction from unstructured text. *International Journal of Geographical Information Science*, 1–20, 747–766. Available from: <https://doi.org/10.1080/13658816.2022.2133125>
- Birch, C.P.D., Oom, S.P. & Beecham, J.A. (2007) Rectangular and hexagonal grids used for observation, experiment and simulation in ecology. *Ecological Modelling*, 206(3), 347–359. Available from: <https://doi.org/10.1016/j.ecolmodel.2007.03.041>
- Borer, M.I. (2006) The location of culture: The urban culturalist perspective. *City & Community*, 5(2), 173–197. Available from: <https://doi.org/10.1111/j.1540-6040.2006.00168.x>
- Calafiore, A., Palmer, G., Comber, S., Arribas-Bel, D. & Singleton, A. (2021) A geographic data science framework for the functional and contextual analysis of human dynamics within global cities. *Computers, Environment and Urban Systems*, 85, 101539. Available from: <https://doi.org/10.1016/j.compenvurbsys.2020.101539>
- Canter, D. (1977) *The psychology of place*. London: Architectural Press.
- Carr, S. & Schissler, D. (1969) The city as a trip: Perceptual selection and memory in the view from the road. *Environment and Behavior*, 1(1), 7–35.
- Chambers, R. (2006) Participatory mapping and geographic information systems: Whose map? Who is empowered and who disempowered? Who gains and who loses? *The Electronic Journal of Information Systems in Developing Countries*, 25(1), 1–11. Available from: <https://doi.org/10.1002/j.1681-4835.2006.tb00163.x>
- Chen, M. (2019) Understanding the dynamics of urban areas of interest through volunteered geographic information. *Journal of Geographical Systems*, 21, 89–109. Available from: <https://doi.org/10.1007/s10109-018-0284-3>
- Corbett, J. & Rambaldi, G. (2009) Geographic information technologies, local knowledge, and change. In: *Qualitative GIS: A mixed methods approach*. London: Sage, pp. 75–92.
- Deacon, B. (2007) County, nation, ethnic group? The shaping of the Cornish identity. *The International Journal of Regional and Local Studies*, 3(1), 5–29. Available from: <https://doi.org/10.1179/jrl.2007.3.1.5>
- Dennett, A. & Wilson, A. (2013) A multilevel spatial interaction modelling framework for estimating interregional migration in Europe. *Environment and Planning A: Economy and Space*, 45(6), 1491–1507. Available from: <https://doi.org/10.1068/a45398>
- Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. (2019) *BERT: Pre-training of deep bidirectional transformers for language understanding*. arXiv:1810.04805 [Cs]. Available from: <https://doi.org/10.48550/arXiv.1810.04805>

- Erlander, S. (1980) *Optimal spatial interaction and the gravity model*. M. Beckmann & H. P. Küenzi, (Eds.). Berlin, Heidelberg: Springer. Available from: <https://doi.org/10.1007/978-3-642-45515-5>
- Fabrikant, S.I., Ruocco, M., Middleton, R., Montello, D.R. & Jörgensen, C. (2002) The first law of cognitive geography: Distance and similarity in semantic space. *Proceedings of GIScience*, 2002, 31–33.
- Farber, S. & Li, X. (2013) Urban sprawl and social interaction potential: An empirical analysis of large metropolitan regions in the United States. *Journal of Transport Geography*, 31, 267–277. Available from: <https://doi.org/10.1016/j.jtrangeo.2013.03.002>
- Farman, J. (2020) *Mobile interface theory: Embodied space and locative media*. New York, NY: Routledge.
- Findlay, A.M., Short, D. & Stockdale, A. (2000) The labour-market impact of migration to rural areas. *Applied Geography*, 20(4), 333–348. Available from: [https://doi.org/10.1016/S0143-6228\(00\)00012-6](https://doi.org/10.1016/S0143-6228(00)00012-6)
- Gao, S., Li, L., Li, W., Janowicz, K. & Zhang, Y. (2017) Constructing gazetteers from volunteered big geo-data based on Hadoop. *Computers, Environment and Urban Systems*, 61, 172–186. Available from: <https://doi.org/10.1016/j.compenvurbsys.2014.02.004>
- Gardner, Z., Mooney, P., De Sabbata, S. & Dowthwaite, L. (2020) Quantifying gendered participation in OpenStreetMap: Responding to theories of female (under) representation in crowdsourced mapping. *GeoJournal*, 85(6), 1603–1620. Available from: <https://doi.org/10.1007/s10708-019-10035-z>
- Glückler, J. & Panitz, R. (2021) Unleashing the potential of relational research: A meta-analysis of network studies in human geography. *Progress in Human Geography*, 45(6), 1531–1557. Available from: <https://doi.org/10.1177/03091325211002916>
- Gong, J., Li, S., Ye, X., Peng, Q. & Kudva, S. (2021) Modelling impacts of high-speed rail on urban interaction with social media in China's mainland. *Geo-spatial Information Science*, 24(4), 638–653. Available from: <https://doi.org/10.1080/10095020.2021.1972771>
- González, M.C., Hidalgo, C.A. & Barabási, A.-L. (2008) Understanding individual human mobility patterns. *Nature*, 453(7196), 779–782. Available from: <https://doi.org/10.1038/nature06958>
- Goodchild, M.F. (2007) Citizens as sensors: The world of volunteered geography. *GeoJournal*, 69(4), 211–221. Available from: <https://doi.org/10.1007/s10708-007-9111-y>
- Goodchild, M.F. & Hill, L.L. (2008) Introduction to digital gazetteer research. *International Journal of Geographical Information Science*, 22(10), 1039–1044. Available from: <https://doi.org/10.1080/13658810701850497>
- Goodchild, M.F. & Li, L. (2012) Assuring the quality of volunteered geographic information. *Spatial Statistics*, 1, 110–120. Available from: <https://doi.org/10.1016/j.spasta.2012.03.002>
- Goodey, B. (1974) *Images of place: Essays on environmental perception, communications and education*. Birmingham: Centre for Urban and Regional Studies, University of Birmingham.
- Gould, P.R. & White, R. (1986) *Mental maps*. Hoboken, NJ: Taylor and Francis.
- Gould, P.R. & White, R.R. (1968) The mental maps of British school leavers. *Regional Studies*, 2(2), 161–182. Available from: <https://doi.org/10.1080/09595236800185171>
- Graham, M., Straumann, R.K. & Hogan, B. (2015) Digital divisions of labor and informational magnetism: Mapping participation in Wikipedia. *Annals of the Association of American Geographers*, 105(6), 1158–1178. Available from: <https://doi.org/10.1080/00045608.2015.1072791>
- Greenberg Raanan, M. & Shoval, N. (2014) Mental maps compared to actual spatial behavior using GPS data: A new method for investigating segregation in cities. *Cities*, 36, 28–40. Available from: <https://doi.org/10.1016/j.cities.2013.09.003>
- Gritta, M., Pilehvar, M.T. & Collier, N. (2020) A pragmatic guide to geoparsing evaluation: Toponyms, named entity recognition and pragmatics. *Language Resources and Evaluation*, 54(3), 683–712. Available from: <https://doi.org/10.1007/s10579-019-09475-3>
- Haesly, R. (2005) Identifying Scotland and Wales: Types of Scottish and Welsh national identities. *Nations and Nationalism*, 11(2), 243–263. Available from: <https://doi.org/10.1111/j.1354-5078.2005.00202.x>
- Haklay, M.E. (2016) *Why is participation inequality important?* London, UK: Ubiquity Press.
- Haklay, M.E. & Weber, P. (2008) Openstreetmap: User-generated street maps. *IEEE Pervasive Computing*, 7(4), 12–18. Available from: <https://doi.org/10.1109/MPRV.2008.80>
- Halterman, A. (2017) Mordecai: Full text geoparsing and event geocoding. *The Journal of Open Source Software*, 2(9), 91. Available from: <https://doi.org/10.21105/joss.00091>
- Han, J., Sun, A., Cong, G., Zhao, W.X., Ji, Z. & Phan, M.C. (2018) Linking fine-grained locations in user comments. *IEEE Transactions on Knowledge and Data Engineering*, 30(1), 59–72. Available from: <https://doi.org/10.1109/TKDE.2017.2758780>
- Haney, W.G. & Knowles, E.S. (1978) Perception of neighborhoods by city and suburban residents. *Human Ecology*, 6(2), 201–214. Available from: <https://doi.org/10.1007/BF00889095>
- Haynes, K.E. & Fotheringham, A.S. (1985) *Gravity and spatial interaction models*. Morgantown: Sage.
- Hecht, B. & Stephens, M. (2014) *A tale of cities: Urban biases in volunteered geographic information*. 9. Available from: <https://doi.org/10.1609/icwsm.v8i1.14554>
- Hollenstein, L. (2008) *Capturing vernacular geography from georeferenced tags* [MSc Thesis]: University of Zurich.
- Hu, Y., Ye, X. & Shaw, S.-L. (2017) Extracting and analyzing semantic relatedness between cities using news articles. *International Journal of Geographical Information Science*, 31(12), 2427–2451. Available from: <https://doi.org/10.1080/13658816.2017.1367797>
- Jang, K.M. & Kim, Y. (2019) Crowd-sourced cognitive mapping: A new way of displaying people's cognitive perception of urban space. *PLoS One*, 14(6), e0218590. Available from: <https://doi.org/10.1371/journal.pone.0218590>
- Kamalloo, E. & Rafiei, D. (2018) *A coherent unsupervised model for toponym resolution*. Proceedings of the 2018 world wide web conference on world wide web - WWW '18, 1287–1296. Lyon, France: ACM Press. Available from: <https://doi.org/10.1145/3178876.3186027>

- Kaplan, S. (1976) Adaptation, structure and knowledge. In: *Environmental knowing*. Stroudsburg, Penn: Downden, Hutchinson and Ross, pp. 32–46.
- Karimzadeh, M., Pezanowski, S., MacEachren, A.M. & Wallgrün, J.O. (2019) GeoTxt: A scalable geoparsing system for unstructured text geo-location. *Transactions in GIS*, 23(1), 118–136. Available from: <https://doi.org/10.1111/tgis.12510>
- Langton, S.H. & Solymosi, R. (2021) Cartograms, hexograms and regular grids: Minimising misrepresentation in spatial data visualisations. *Environment and Planning B: Urban Analytics and City Science*, 48(2), 348–357. Available from: <https://doi.org/10.1177/2399808319873923>
- Laurier, E., Brown, B. & McGregor, M. (2016) Mediated pedestrian mobility: Walking and the map app. *Mobilities*, 11(1), 117–134. Available from: <https://doi.org/10.1080/17450101.2015.1099900>
- Lee, T.R. (1973) Psychology and living space. In: Downs, R.M. & Stea, D. (Eds.) *Image and environment: Cognitive mapping and spatial behavior*. Chicago, IL: Aldine, pp. 87–108.
- Leidner, J.L. & Lieberman, M.D. (2011) Detecting geographical references in the form of place names and associated spatial natural language. *SIGSPATIAL Special*, 3(2), 5–11. Available from: <https://doi.org/10.1145/2047296.2047298>
- Li, C. & Sun, A. (2014) *Fine-grained location extraction from tweets with temporal awareness*. Proceedings of the 37th international ACM SIGIR conference on Research & Development in information retrieval - SIGIR '14, July 6–11, 2014, pp. 43–52. Gold Coast, Queensland, Australia, ACM Press. Available from: <https://doi.org/10.1145/2600428.2609582>
- Li, Z., Huang, X., Ye, X., Jiang, Y., Martin, Y., Ning, H. et al. (2021) Measuring global multi-scale place connectivity using geotagged social media data. *Scientific Reports*, 11(1), 14694. Available from: <https://doi.org/10.1038/s41598-021-94300-7>
- Lin, J., Wu, Z. & Li, X. (2019) Measuring inter-city connectivity in an urban agglomeration based on multi-source data. *International Journal of Geographical Information Science*, 33(5), 1062–1081. Available from: <https://doi.org/10.1080/13658816.2018.1563302>
- Liu, Y., Wang, F., Kang, C., Gao, Y. & Lu, Y. (2014) Analyzing relatedness by toponym Co-occurrences on web pages: Analyzing relatedness by toponym Co-occurrences on web pages. *Transactions in GIS*, 18(1), 89–107. Available from: <https://doi.org/10.1111/tgis.12023>
- Lukermann, F. (1961) The concept of location in classical geography. *Annals of the Association of American Geographers*, 51(2), 194–210. Available from: <https://doi.org/10.1111/j.1467-8306.1961.tb00373.x>
- Lynch, K. (1964) *The image of the city*. Cambridge, MA: MIT Press.
- Marston, S.A., Jones, J.P., III & Woodward, K. (2005) Human geography without scale. *Transactions of the Institute of British Geographers*, 30(4), 416–432. Available from: <https://doi.org/10.1111/j.1475-5661.2005.00180.x>
- Massey, D. (2008) A global sense of place. In: *The cultural geography reader*. London: Routledge, pp. 269–275.
- Medvedev, A.N., Lambiotte, R. & Delvenne, J.-C. (2019) The anatomy of reddit: An overview of academic research. In: Ghanbarnejad, F., Saha Roy, R., Karimi, F., Delvenne, J.-C. & Mitra, B. (Eds.) *Dynamics on and of complex networks III*. Cham, Switzerland: Springer International Publishing, pp. 183–204. Available from: [https://doi.org/10.1007/978-3-030-14683-2\\_9](https://doi.org/10.1007/978-3-030-14683-2_9)
- Meijers, E. & Peris, A. (2019) Using toponym co-occurrences to measure relationships between places: Review, application and evaluation. *International Journal of Urban Sciences*, 23(2), 246–268. Available from: <https://doi.org/10.1080/12265934.2018.1497526>
- Merrifield, A. (1993) Place and space: A Lefebvrian reconciliation. *Transactions of the Institute of British Geographers*, 18(4), 516–531. Available from: <https://doi.org/10.2307/622564>
- Merriman, P. (2012) Human geography without time-space. *Transactions of the Institute of British Geographers*, 37(1), 13–27. Available from: <https://doi.org/10.1111/j.1475-5661.2011.00455.x>
- Middleton, C. & Freestone, P. (2008) *The impact of culture-led regeneration on regional identity in north East England*. Regional and Urban Regeneration in European Peripheries, Bratislava: Institute of Public Policy, pp. 51–58.
- Miller, H.J. (2018) Time geography. In: *Handbook of behavioral and cognitive geography*. Cheltenham, UK; Northampton, MA: Edward Elgar Publishing, pp. 74–94. Available from: <https://doi.org/10.4337/9781784717544.00011>
- Miyoshi, C. & Givoni, M. (2013) The environmental case for the high-speed train in the UK: Examining the London. *International Journal of Sustainable Transportation*, 8(2), 107–126. Available from: <https://doi.org/10.1080/15568318.2011.645124>
- Moncla, L., Renteria-Agualimpia, W., Nogueras-Iso, J. & Gaio, M. (2014) *Geocoding for texts with fine-grain toponyms: An experiment on a geoparsed hiking descriptions corpus*. Proceedings of the 22nd ACM SIGSPATIAL international conference on advances in geographic information systems, 183–192. Dallas, TX: ACM. Available from: <https://doi.org/10.1145/2666310.2666386>
- Montello, D.R. (1993) Scale and multiple psychologies of space. In: Goos, G., Hartmanis, J., Frank, A.U. & Campari, I. (Eds.) *Spatial information theory a theoretical basis for GIS* (Vol. 716). Berlin, Heidelberg, Germany: Springer Berlin Heidelberg, pp. 312–321. Available from: [https://doi.org/10.1007/3-540-57207-4\\_21](https://doi.org/10.1007/3-540-57207-4_21)
- Montello, D.R., Goodchild, M.F., Gottsegen, J. & Fohl, P. (2003) *Where's Downtown?: Behavioral methods for determining referents of vague spatial queries*. 20.
- Moseley, M.J. (2023) *Accessibility: The rural challenge*. Abingdon: Taylor & Francis.
- Murphy, N. (2019) Reddit's 2019 year in review - Upvoted. Available from: <https://www.redditinc.com/blog/reddits-2019-year-in-review/#content> [Accessed 27th April 2022].
- Murray, D. & Spencer, C. (1979) Individual differences in the drawing of cognitive maps: The effects of geographical mobility, strength of mental imagery and basic graphic ability. *Transactions of the Institute of British Geographers*, 4(3), 385. Available from: <https://doi.org/10.2307/622058>
- Nayak, A. (2016) *Race, place and globalization: Youth cultures in a changing world*. London, UK: Bloomsbury Publishing.
- Noulas, A., Scellato, S., Mascolo, C. & Pontil, M. (2011) An empirical study of geographic user activity patterns in foursquare. *Proceedings of the International AAAI Conference on Web and Social Media*, 5(1), 570–573. Available from: <https://doi.org/10.1609/icwsm.v5i1.14175>



- Office for National Statistics. (2022) *How the population changed in Newcastle upon Tyne, Census 2021* - ONS. Available from: <https://www.ons.gov.uk/visualisations/censuspopulationchange/E08000021/> [Accessed 17th September 2023].
- Ostermann, F.O., Huang, H., Andrienko, G., Andrienko, N., Capineri, C., Farkas, K. et al. (2015) Extracting and comparing places using Geo-social Media. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, II-3/W5, 311–316. Available from: <https://doi.org/10.5194/isprsannals-II-3-W5-311-2015>
- Pánek, J. (2016) From mental maps to GeoParticipation. *The Cartographic Journal*, 53(4), 300–307. Available from: <https://doi.org/10.1080/00087041.2016.1243862>
- Patterson, Z. & Farber, S. (2015) Potential path areas and activity spaces in application: A review. *Transport Reviews*, 35(6), 679–700. Available from: <https://doi.org/10.1080/01441647.2015.1042944>
- Peake, S. & Moore, T. (2004) *Analysis of distortions in a mental map using GPS and GIS*. 10.
- Pierce, J., Martin, D.G. & Murphy, J.T. (2011) Relational place-making: The networked politics of place. *Transactions of the Institute of British Geographers*, 36(1), 54–70. Available from: <https://doi.org/10.1111/j.1475-5661.2010.00411.x>
- Pocock, D.C.D. (1976) Some characteristics of mental maps: An empirical study. *Transactions of the Institute of British Geographers*, 1(4), 493. Available from: <https://doi.org/10.2307/621905>
- Purves, R.S., Clough, P., Jones, C.B., Hall, M.H. & Murdock, V. (2018) *Geographic information retrieval: Progress and challenges in spatial search of text*. Foundations and Trends in Information Retrieval, 12(2-3), pp. 164–318. Available from: <https://doi.org/10.1561/15000000034>
- Rae, A. (2009) From spatial interaction data to spatial interaction information? Geovisualisation and spatial structures of migration from the 2001 UK census. *Computers, Environment and Urban Systems*, 33(3), 161–178. Available from: <https://doi.org/10.1016/j.compenvurbsys.2009.01.007>
- Rey, S., Arribas-Bel, D. & Wolf, L.J. (2023) *Geographic data science with python*. Boca Raton, FL: CRC Press.
- Roos Breines, M., Raghuram, P. & Gunter, A. (2019) Infrastructures of immobility: Enabling international distance education students in Africa to not move. *Mobilities*, 14(4), 484–499. Available from: <https://doi.org/10.1080/17450101.2019.1618565>
- Rowe, F., Calafiore, A., Arribas-Bel, D., Samardzhiev, K. & Fleischmann, M. (2022) Urban exodus? Understanding human mobility in Britain during the COVID-19 pandemic using meta-Facebook data. *Population, Space and Place*, 29(1), e2637. Available from: <https://doi.org/10.1002/psp.2637>
- Rowe, F., Lovelace, R. & Dennett, A. (2022) Spatial interaction modelling: A Manifesto. OSF Preprints. Available from: <https://doi.org/10.31219/osf.io/xcdms>
- SafeGraph. (2022) *Places data curated for accurate geospatial analytics*|SafeGraph. Available from: <https://www.safegraph.com> [Accessed 27th June 2022].
- Salvini, M.M. & Fabrikant, S.I. (2016) Spatialization of user-generated content to uncover the multirelational world city network. *Environment and Planning, B, Planning & Design*, 43(1), 228–248. Available from: <https://doi.org/10.1177/0265813515603868>
- Shaw, J. & Hesse, M. (2010) Transport, geography and the “new” mobilities. *Transactions of the Institute of British Geographers*, 35(3), 305–312. Available from: <https://doi.org/10.1111/j.1475-5661.2010.00382.x>
- Singleton, A.D., Wilson, A.G. & O'Brien, O. (2012) Geodemographics and spatial interaction: An integrated model for higher education. *Journal of Geographical Systems*, 14(2), 223–241. Available from: <https://doi.org/10.1007/s10109-010-0141-5>
- Smith, T.A., Laurier, E., Reeves, S. & Dunkley, R.A. (2020) “Off the beaten map”: Navigating with digital maps on moorland. *Transactions of the Institute of British Geographers*, 45(1), 223–240. Available from: <https://doi.org/10.1111/tran.12336>
- Statista. (2022) *Most popular social networks worldwide as of January 2022, ranked by number of monthly active users*. Available from: <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/> [Accessed 11th July 2022].
- Steiger, E., Westerholt, R., Resch, B. & Zipf, A. (2015) Twitter as an indicator for whereabouts of people? Correlating twitter with UK census data. *Computers, Environment and Urban Systems*, 54, 255–265. Available from: <https://doi.org/10.1016/j.compenvurbsys.2015.09.007>
- Stock, K., Pasley, R.C., Gardner, Z., Brindley, P., Morley, J. & Cialone, C. (2013) Creating a corpus of geospatial natural language. In: Hutchison, D., Kanade, T., Kittler, J., Kleinberg, J.M., Mattern, F., Mitchell, J.C. et al. (Eds.) *Spatial information theory* (Vol. 8116. Cham, Switzerland: Springer International Publishing, pp. 279–298. Available from: [https://doi.org/10.1007/978-3-319-01790-7\\_16](https://doi.org/10.1007/978-3-319-01790-7_16)
- Sylla, A., Glawe, F., Braun, D., Padev, M., Schäfer, S., Ahmetaj, A. et al. (2022) Discourses of climate delay in American reddit discussions. In: Spezzano, F., Amaral, A., Ceolin, D., Fazio, L. & Serra, E. (Eds.) *Disinformation in open online media* (Vol. 13545. Cham, Switzerland: Springer International Publishing, pp. 123–137. Available from: [https://doi.org/10.1007/978-3-031-18253-2\\_9](https://doi.org/10.1007/978-3-031-18253-2_9)
- Taaffe, E.J., Gauthier, H.L. & O'Kelly, M.E. (1996) *Geography of transportation*. Upper Saddle River: Prentice Hall.
- Taylor, P.J. (1983) *Distance decay in spatial interactions*. Concepts and Techniques in Modern Geography, Norwich: Geo Books, p. 2.
- Thomas, R.W. (1981) *Information statistics in geography*. Norwich: Geo Abstracts.
- Titheridge, H., Achuthan, K., Mackett, R.L. & Solomon, J. (2009) Assessing the extent of transport social exclusion among the elderly. *Journal of Transport and Land Use*, 2(2), 31–48. Available from: <https://doi.org/10.5198/jtlu.v2i2.44>
- Tobler, W.R. (1970) A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46, 234. Available from: <https://doi.org/10.2307/143141>
- Won, M., Murrieta-Flores, P. & Martins, B. (2018) Ensemble named entity recognition (NER): Evaluating NER tools in the identification of place names in historical corpora. *Frontiers in Digital Humanities*, 5, 2. Available from: <https://doi.org/10.3389/fdigh.2018.00002>
- Worboys, M.F. (2001) Nearness relations in environmental space. *International Journal of Geographical Information Science*, 15(7), 633–651. Available from: <https://doi.org/10.1080/13658810110061162>

- Yang, Y., Li, D. & Li (Robert), X. (2019) Public transport connectivity and intercity tourist flows. *Journal of Travel Research*, 58(1), 25–41. Available from: <https://doi.org/10.1177/0047287517741997>
- Ye, X., Gong, J. & Li, S. (2021) Analyzing Asymmetric City connectivity by toponym on social Media in China. *Chinese Geographical Science*, 31(1), 14–26. Available from: <https://doi.org/10.1007/s11769-020-1172-6>
- Yin, W., Kann, K., Yu, M. & Schütze, H. (2017) *Comparative study of CNN and RNN for natural language processing*. arXiv:1702.01923 [Cs]. Available from: <https://doi.org/10.48550/arXiv.1702.01923>
- Zook, M. (2006) The geographies of the internet. *Annual Review of Information Science and Technology*, 40(1), 53–78. Available from: <https://doi.org/10.1002/aris.144040s>

**How to cite this article:** Berragan, C., Singleton, A., Calafiore, A. & Morley, J. (2024) Mapping cognitive place associations within the United Kingdom through online discussion on Reddit. *Transactions of the Institute of British Geographers*, 49, e12669. Available from: <https://doi.org/10.1111/tran.12669>