

Mapping Cognitive Place Connectivity within the United Kingdom through Online Discussion on Reddit

Cillian Berragan, Alex Singleton, Alessia Calafiore & Jeremy Morley

Geographic Data Science Lab

University of Liverpool



Motivation

What is Cognitive Place Connectivity?

- ▶ Traditional connectivity between places typically explored through true population movements with both **temporal** and **spatial** dimension
- ▶ Captures persistent connections between distinct locations; **not limited** temporally or spatially, and generated from **mental maps**

What are Mental Maps?

- ▶ The **cognitive visualisation** of a geographic environment
- ▶ Represent collective, **experiential geographic knowledge**, relating to both places, and the relationships between them
- ▶ Exhibit **biases**:
 - ▶ More detailed with respect to locations that we are familiar with
 - ▶ Prominent features may be perceived as larger; e.g. city skylines or busy roads
 - ▶ Perceived distance between locations influenced by travel time

Quantifying Cognitive Place Connectivity

▶ **Social Media text**

- ▶ Captures informal **experiential geographic knowledge**, rather than structured volunteered data like Open Street Map
- ▶ **Passively contributed** by many individuals in a **large volume** online (unlike surveys)
- ▶ Contains **embedded geographic information** through place names

▶ **Co-occurring** place names in a shared textual context have implicit perceived connectivity

- ▶ Demonstrates a general association between locations built from a variety contexts
- ▶ More co-occurrences leads to higher cognitive connectivity

Aims

Distance decay can be assumed when considering physical movements, but how does distance affect cognitive connections?

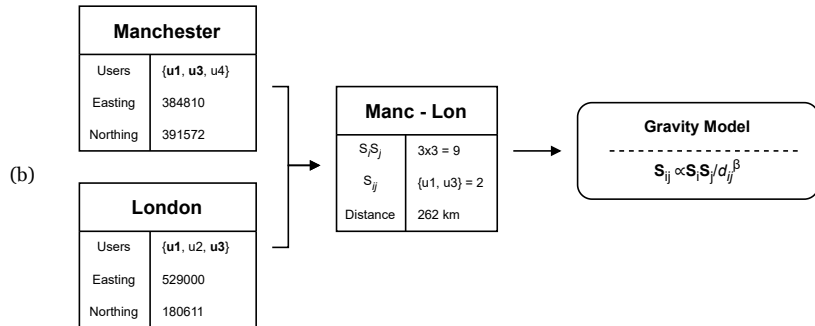
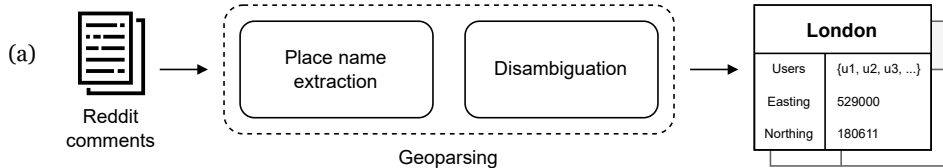
Aim 1: Quantify cognitive place connections across the UK using Reddit comments

Aim 2: Evaluate the effect of distance on the strength of these connections

Data

- ▶ Extracted **8.3 million** comments submitted by **500 thousand** unique users from Reddit's UK related subreddits to build cognitive place connections
- ▶ NER model trained using:
 - ▶ **WNUT17**: 5,690 comments from Reddit, YouTube, Twitter, StackExchange; kept only *LOC* entities
 - ▶ **Reddit**: 498 comments from UK subreddits, annotated with place names
- ▶ Why use Reddit?
 - ▶ More focus on the discussion of places in chosen subreddits
 - ▶ Longer form text with a more structured written style
 - ▶ Accessible API & Pushshift archive

Methods



Methods

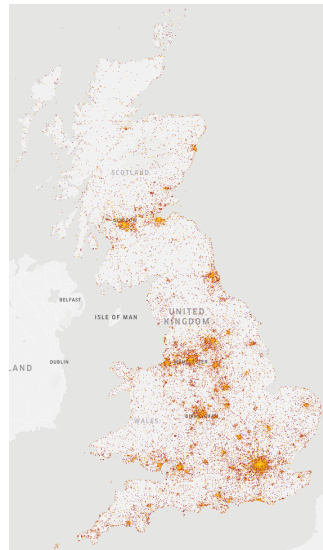
- ▶ Used a custom **geoparsing** pipeline to identify all place names, and ground to geographic locations
 - ▶ *NER stage*: Fine-tuned **transformer** model, targeted specifically towards comments on Reddit; Reddit/WNUT training data
 - ▶ *Disambiguation stage*: Used OS Open Names and the Gazetteer of British Place Names; includes individual streets and some points of interest
- ▶ **26.1%** of comments contained a place name; **4.6 million** in total
 - ▶ **15 million** unique cognitive connections
- ▶ Generate a gravity model to explore the distance decay on these connections;

$$\log(\mathbf{S}_{ij}) = b_0 + b_1 \log(\mathbf{S}_i \mathbf{S}_j) + b_2 \log(d_{ij}) \quad (1)$$

$$\beta = -\frac{b_2}{b_1} \quad (2)$$

Results; Geoparsing

- ▶ NER model performance (just WNUT data): **F_1 : 0.72**, Recall: 0.81, Precision: 0.71 (498 Reddit comments)
 - ▶ Higher than WNUT test performance; reflects Reddits more structured written style
- ▶ Toponym disambiguation: **63%** of all place names were attributed with coordinates
- ▶ **54,215** unique locations; many have shared place names
- ▶ 1% of users contribute 32% of all place names (top 2,079 users)



Results; Distance decay

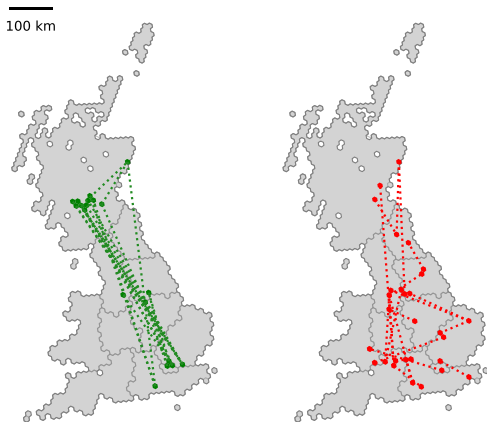


Figure 1: Top and bottom linear model residuals

- ▶ Aggregated to **H3 polygons**; reduces noise, combines connections
- ▶ β coefficient of **1.05**; higher than past work that only considers cities
- ▶ Regression model **R^2 of 0.754**; some regions do not conform with the global β coefficient
- ▶ **High residuals** show connections **stronger** than expected
 - ▶ Largest: London - Edinburgh (3.54), Glasgow - London (3.47)
- ▶ **Low residuals** show connections **weaker** than expected (Red)

Results; Mixed effects

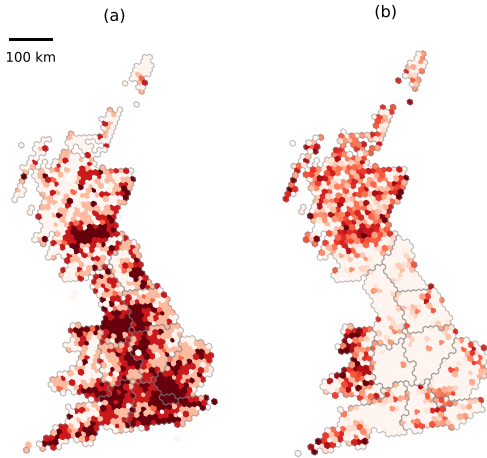


Figure 2: (a) Distribution of mentioned locations. (b) β coefficients for mixed model.

- ▶ Calculated **Mixed Effects Model** allowing intercept and slope for distance to vary across regions
- ▶ Strong distance decay for rural Wales, rural Scotland, but not true for rural England
- ▶ Urban Scotland, surrounding London and the South show strong distance decay, not found in other urban areas like the Midlands and North of England
- ▶ Suggests population or popularity of locations is not the only factor affecting the observed levels of cognitive distance decay

Results; Mapping distance decay

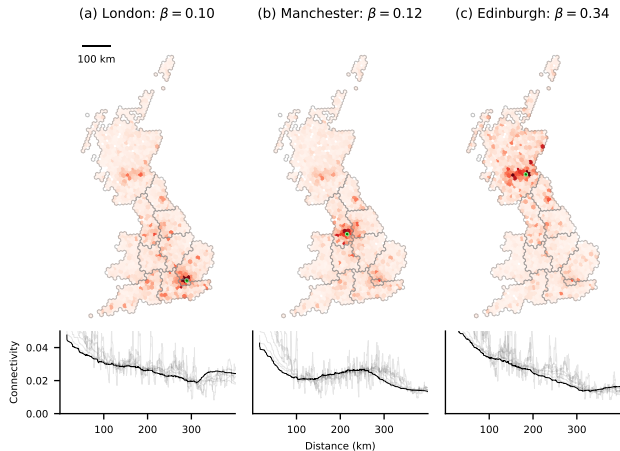


Figure 3: Connection strength, β coefficients and decay curves.

- ▶ Varying patterns in distance decay
- ▶ London lowest β coefficient; least sharp decay curve, small increase when reaching urban areas in Scotland
- ▶ Manchester similar β but different pattern; sharp drop initially, which increases in range of London or Edinburgh
- ▶ Edinburgh distinct; steeper curve, only increasing within range of London. Largely influenced by strong connections below 100km in Scotland

Conclusion


- ▶ Geographic information may be parsed from Reddit comments, despite no explicit geographic information (e.g. geotags)
- ▶ Captures unique patterns in cognitive distance decay between locations
 - ▶ Largely influenced by proximity to major cities
 - ▶ Patterns are unique to each city

Future work

- ▶ Different locations may share varying similarities in their semantic typology
 - ▶ Clustering through LDA
 - ▶ Cosine similarity of lexicons
- ▶ Locations mentioned may give insight into urban areas of interest, or semantic regions
- ▶ Exploring difference between Reddit communities (subreddits)

Thank you for listening!

 GitHub: [cjber/reddit-connectivity](https://github.com/cjber/reddit-connectivity)

 Twitter: [@cjberragan](https://twitter.com/cjberragan)

 Email: cjberragan@gmail.com

