Spatial-Temporal Patterns in Twitter Activity Related to 'Fracking' in Great Britain

Bartie P[*1], Dickie J[2], Varley A[2], Ryder S[3],
Evensen D[4], Devine-Wright P[3], Whitmarsh L[5]

[1] Computer Science, Heriot-Watt University
[2] Biology and Environmental Sciences, University of Stirling
[3] Geography, University of Exeter
[4] Politics and International Relations, University of Edinburgh
[5] Psychology, University of Bath

February 12, 2021

Summary

This research explores the spatial variation of 317 million UK based geotweets as a measure of the level online public engagement with shale gas extraction ('fracking') in Great Britain. Fracking has proven to be a contentious issue and has received wide news coverage. This analysis was carried out to compare the spatial distribution of online activity related to shale gas exploration sites. To normalise the results for population density a χ-squared expectation surface was calculated revealing higher than expected levels of interested near the active fracking site of Preston New Road and licenced extraction blocks in Lancashire. The level of sustained engagement was measured by time slicing the map, as well monitoring distinct Twitter accounts to reduce the impact of bots and over enthusiastic users. The conclusion is that spatial proximity of energy infrastructure projects does appear to play a part in the patterns of online engagement.

KEYWORDS: PostgreSQL, fracking, geolocated tweets, χ-squared expectation surface, social media

1. Introduction

Hydrocarbon based energy sources such as coal, oil, and gas, have fuelled growth in the developed world for decades. To increase the supply of fossil fuels sourced from the UK the option of hydraulic fracturing, also known as 'fracking', has been explored in recent years. This technique involves pumping sand and chemicals into porous shale to force trapped natural gas to flow into drilled wells (Speight, 2013). The UK controls this through issuing Petroleum Exploration and Development Licences (PEDL) (Cotton, 2017). The topic is contentious as it involves horizontal drilling that can go under nearby buildings and can cause small earth tremors.

This research looks at Great Britain's (GB) public engagement with unconventional hydrocarbons (i.e. fracking) based on 317 million geotweets collected over a 5 year period from 2015 to 2020. Our findings show that there are spatial and temporal patterns in the online social engagement with 'fracking' related topics, both environmentally and politically. Twitter activity was normalised using a χ-squared expectation surface, which shows that higher than expected levels of online engagement exist near the UK's only active fracking site (Preston New Road in Lancashire). Peak activity is noted during

[*] phil.bartie@hw.ac.uk

politically historical decisions such as the moratoriums in Scotland (2017) and England (2019), and the key events such as UK's largest fracking induced tremor (Aug 2019).

Social media has seen one of the fastest adoptions of any online platform, with an expected audience of 3.43 billion users by 2023 Twitter (Clement, 2020), and is one of the most dominant micro-blogging and social networking platforms with over 330 million active users each month (Clement, 2019). It has revolutionised our ability to reach an audience, both in the rate and range of information dispersion. According to Statistica there are around 17 million Twitter users based in the UK, of which 43% use Twitter several times per day, and the most followed Twitter account in the UK is the BBC Breaking News account (Tankovska, 2021).

2. Data Collection and Processing

Twitter offers the facility for software developers to connect to various data streams via an Application Programming Interface (API) to request subsets of their data. As a result researchers have been able to analyse tweets to study online movements, such as the Arab Spring uprisings (Meraz and Papacharissi, 2013), and anti-fracking groups (Hopke, 2015). The Streaming API makes a sample of tweets available in real time, filtered either by keywords or location. For this research only tweets that have a location attribute along with the message, username, language, and timestamp were used. These geotweets make up only a small fraction of all tweets but appear to not be filtered by Twitter, meaning we are collecting most if not all of the geotweets from GB during the study period. The Tweepy Python library was used to connect to the Twitter Streaming API with a location based filter for GB, collecting on average 200,000 tweets per day covering a wide range of subjects, of which only a small fraction were be related to fracking.

2.1. Database Tuning

The PostgreSQL v10 database server was configured to maximise performance, enabling parallel workers, maximising memory limits and moving the WAL (Write Ahead Log) to a 1TB NVME drive with a write performance of 3500MB/s. The data tables were distributed across three tablespaces each on a separate physical disks. As well as spatial and date indexes, a `gin` index was used with `trigrams` to support partial matching of Tweet messages (as below), which greatly improved search performance.

```
CREATE INDEX idx_msg ON tweets USING GIN(msg gin_trgm_ops);
```

3. Results

3.1. Spatial Pattern of UK Geotweets

The overall spatial pattern of GB geotweets reflects the population density, with cities and urban regions creating the majority of messages, as shown in the aggregated counts per 10km by 10km grid cell in Figure 1.

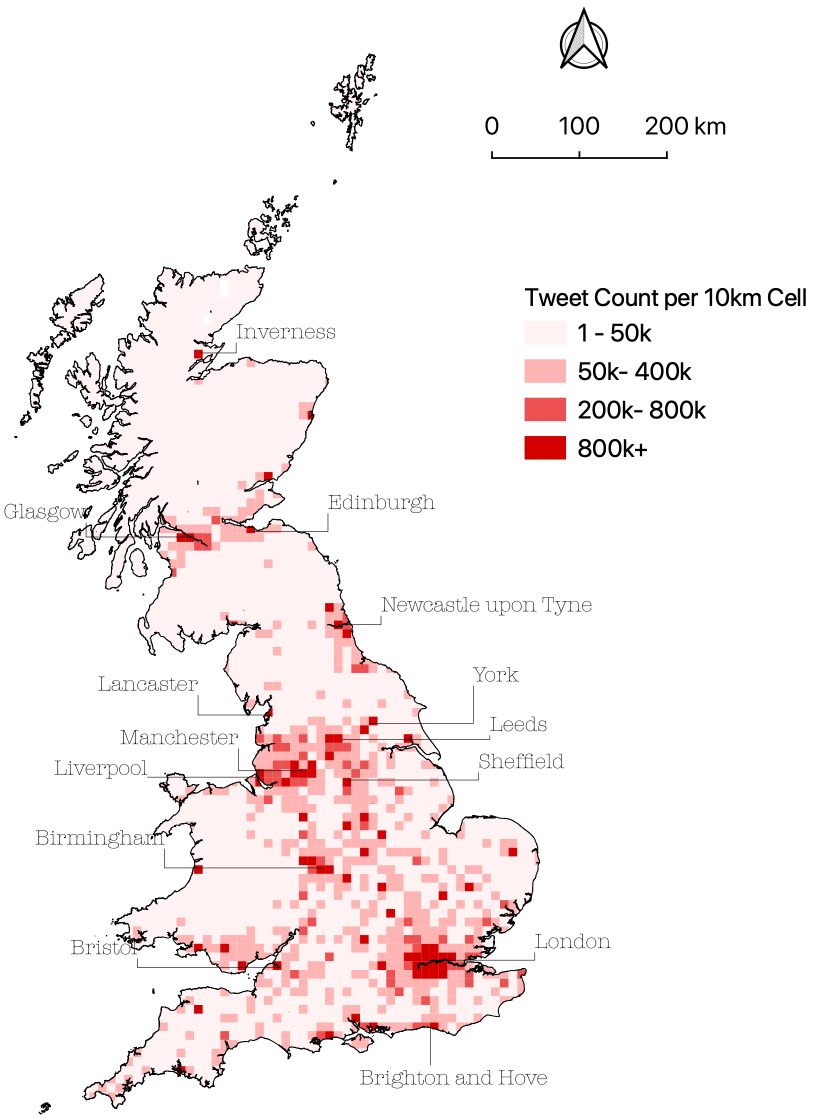

Figure 1: Twitter activity from 2015 to 2020 (all geolocated tweets)

For the period from 3 March 2015 until 3 March 2020 (5 years) there were 317 million geotweets analysed for GB of which 25,196 related to fracking based on the definition that the message contained 'frack'', 'shale gas', or 'hydraulic frac'. This filter would select messages that contain partial matches, such as 'fracking', 'hydraulic fracturing', 'hydraulic fractured'. Figure 2 shows the largest clusters of 'fracking' messages are sent from London, Liverpool, and York.

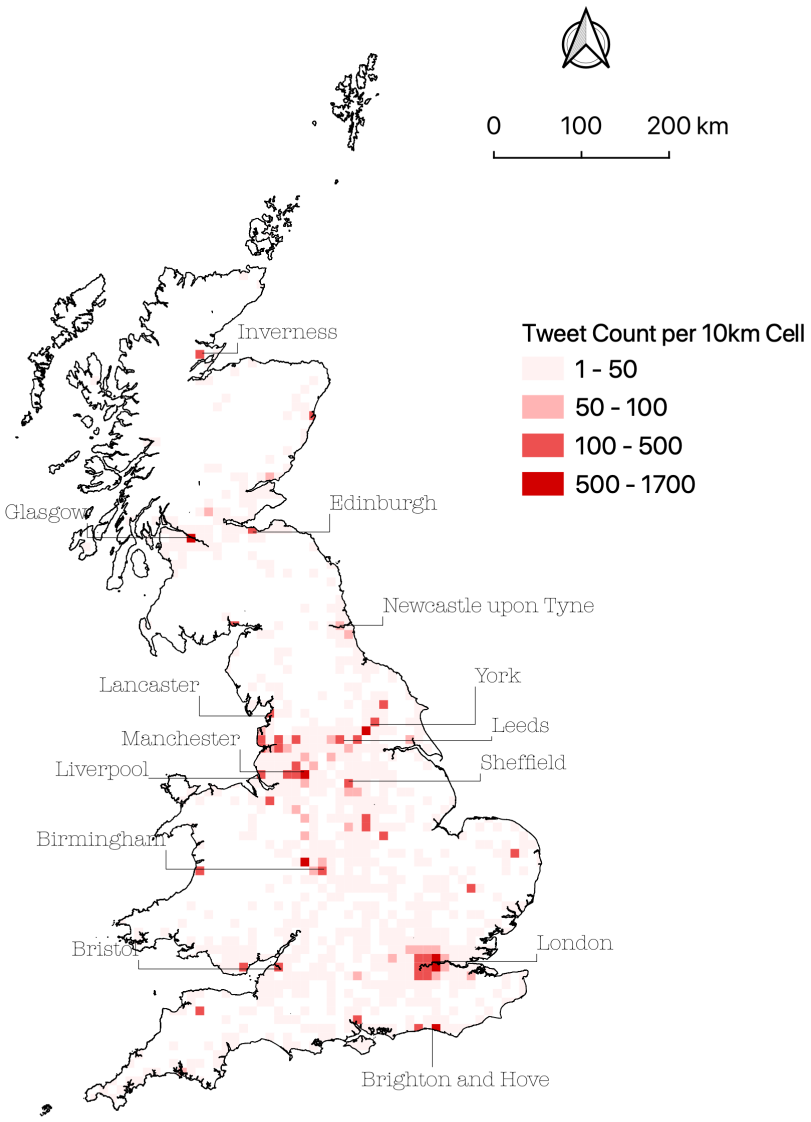

Figure 2: Tweet counts related to 'fracking' from 2015-2020

However, to understand the population's interest and activity in the topic it's necessary to normalise the data as a proportion of all Twitter activity to take account of the variation in the Twitter user density. This was done using a χ-squared expectation surface, based on work by Wartmann and Mackaness (2020), which takes into account the distribution of tweets in each grid cell (exp), and the number of 'fracking' related tweets in each cell (obs).

$$\chi = \frac{(obs - \exp)}{\sqrt{exp}} \tag{1}$$

The calculations indicate greater (positive values), equal (around 0), or fewer (negative values) fracking related tweets than expected. Greater than expected interest in fracking can be noted in Figure 3 around Lancashire, and lower than expected in London, Newcastle, and Birmingham.

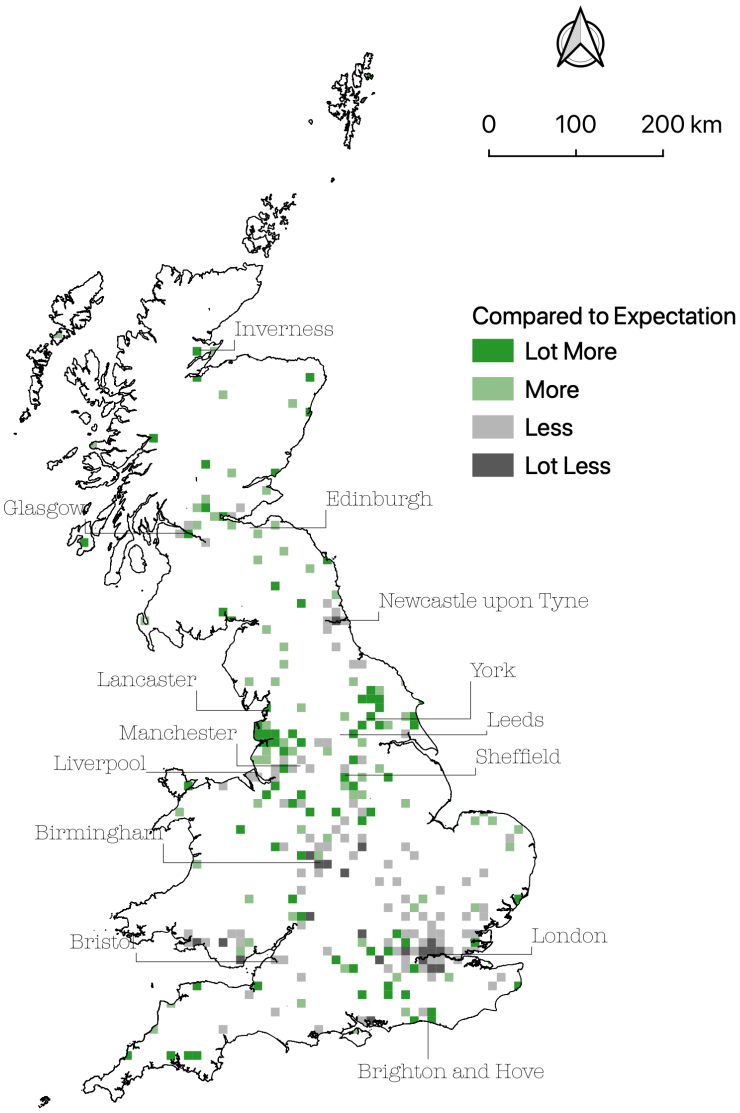

Figure 3: χ-squared surface expectation for fracking related tweet activity - showing regions with higher and lower than expected numbers of tweets

Figure 4 shows the largest 3 clusters of higher than expected activity, based on DBScan (epsilon=20km), overlaid with the PEDL areas. These highly active regions correlate with the PEDL blocks, indicating a greater awareness and interest in the topic from the public near those regions. There are also a number of other active regions away from licences blocks, such as Glasgow, that need further investigation.

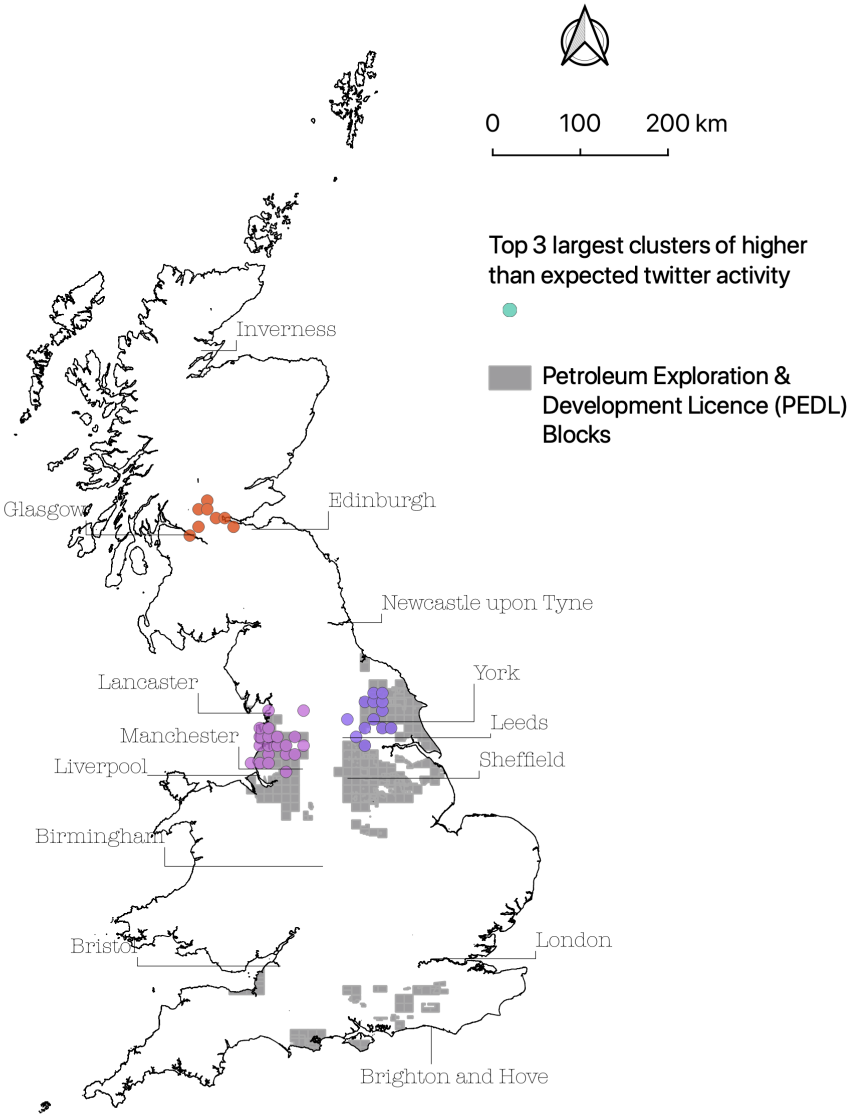

Figure 4: Licenced blocks and three largest clusters of higher than expected Twitter activity

3.2. Sustained Activity

To identify if these active regions have arisen from a short term burst of activity or a sustained interest in fracking related topics the dataset was time-sliced. This was done by creating a moving time window and monitoring the change of activity within each grid cell based on the χ-squared surface expectation model. A 12 month moving window was used, incremented in 1 month steps, with the number of fracking tweets recorded for each cell for each step being compared to expected numbers. The difference between each step was calculated and summed, so that if activity remained higher than expected (χ-squared value) the resulting value would be a high positive value, whilst if the activity fluctuated the resultant value would be around zero, and if the activity was consistently lower than expected the total for that cell would be a large negative value (Fig 5).

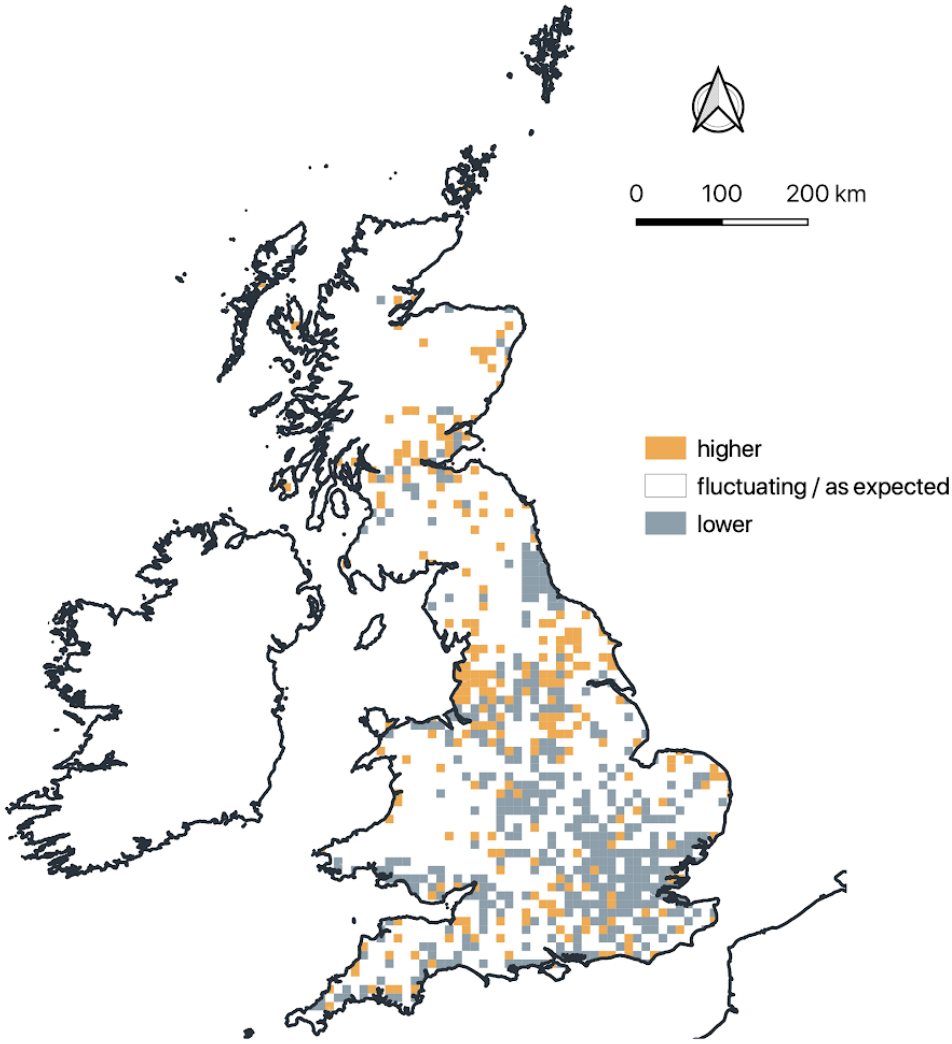

Figure 5: Consistency of fracking related Twitter activity, showing regions that are consistently higher/lower than expected, and those which fluctuate or are as expected in years 2015-2020

This shows a continued lower than expected interest around London and Newcastle, but a higher than expected interest around Lancashire and Liverpool, near Preston New Road's active drilling site.

3.3. Distinct Users

Whilst numbers of tweets indicates the activity level it can be skewed by a few very active users, or bots. One way to avoid such biases is to measure topic interest as calculated by the number of unique usernames tweeting about fracking per 10km by 10km grid cell, thereby limiting an active account to a single input in each cell. However on its own this will just reveal the underlying population density across the country, so the results need to be normalised by the total number of distinct usernames per grid cell. Figure 6 shows the output of that calculation, where the darker regions have a higher ratio of distinct users mentioning fracking related topics as a proportion of the count of all distinct Twitter usernames within the grid cell. Notably the major hotspots are in Scotland (Glasgow, Falkirk/Grangemouth), around the cities of Liverpool, Lancashire, and Sheffield, correlating to the findings in Fig 4.

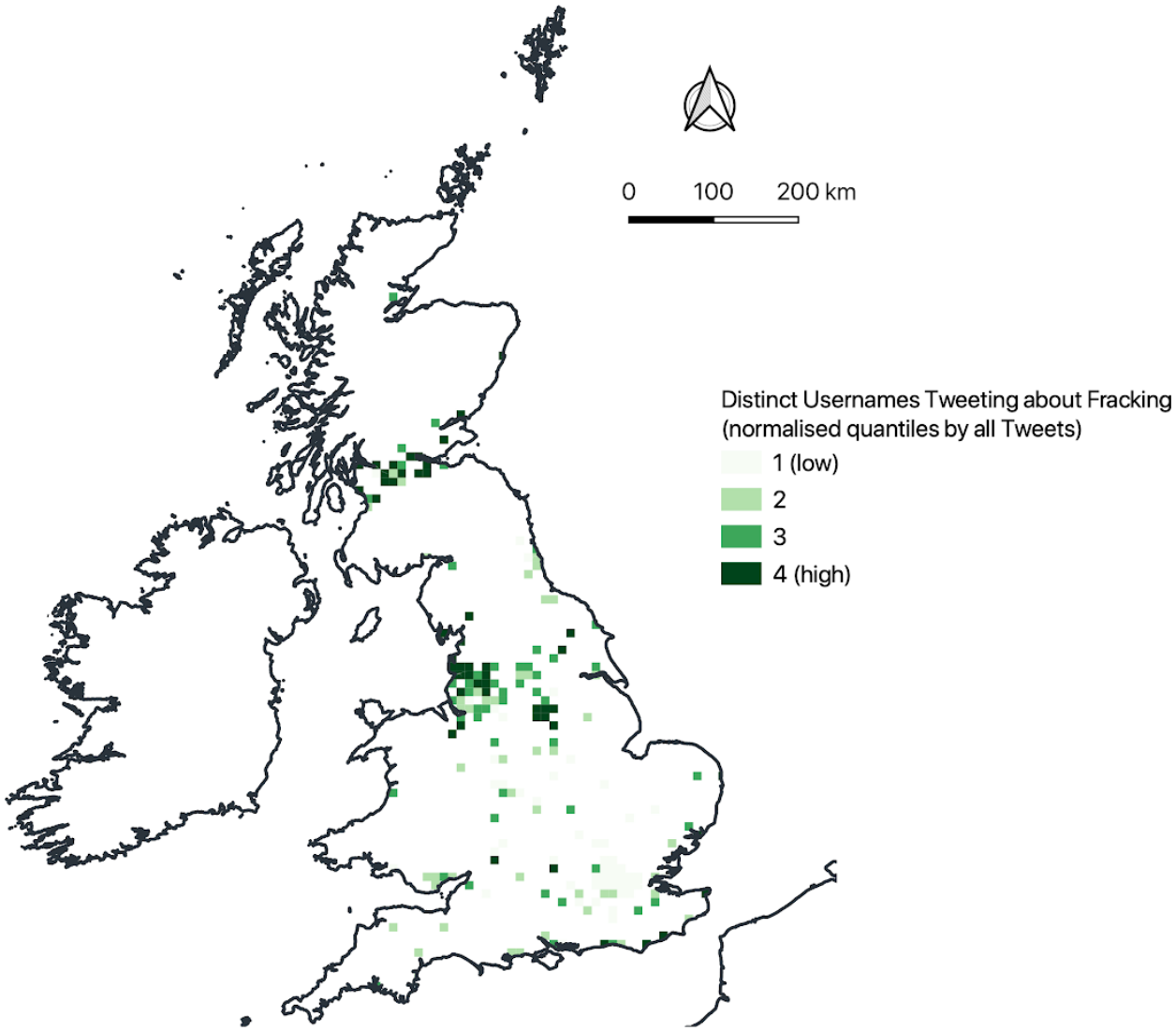

Figure 6: Number of unique users per grid cell over 5 years mentioning fracking keywords normalised by unique usernames from all tweets in the cell (shown as quantiles)

4. Conclusions and Future Work

These results show a spatial pattern of higher engagement relating to 'fracking' than would be expected from the underlying Twitter population in certain regions. These regions coincide with the largest PEDL areas, and only active UK shale gas extraction site. A Scottish cluster can be attributed to political events, based on examining the tweet messages (e.g. Scottish 'fracking' ban). The next phase of this research will be to explore the temporal dimension in greater detail and compare these findings to 3 annual waves of a longitudinal study that has been conducted through YouGov across Great Britain, as part of the ASSIST project.

5. Acknowledgements

This research has been undertaken as part of the 'The Attitudes to Shale GaS in Space and Time' (ASSIST) project funded through ESRC and NERC (Grant Ref: NE/R017727/1). The authors also thank Twitter Inc. for supporting research through their developer programme.

6. References

Clement, J. (2019) *Twitter: number of active users 2010-2019 | Statista*, *Statista*.

Clement, J. (2020) *Number of social network users worldwide from 2010 to 2023*, *Statista*.

Cotton, M. (2017) 'Fair fracking? Ethics and environmental justice in United Kingdom shale gas policy and planning', *Local Environment*. doi: 10.1080/13549839.2016.1186613.

Hopke, J. E. (2015) 'Hashtagging Politics: Transnational Anti-Fracking Movement Twitter Practices', *Social Media and Society*. doi: 10.1177/2056305115605521.

Meraz, S. and Papacharissi, Z. (2013) 'Networked Gatekeeping and Networked Framing on #Egypt', *International Journal of Press/Politics*. doi: 10.1177/1940161212474472.

Speight, J. G. (2013) *Shale Gas Production Processes*, *Shale Gas Production Processes*. doi: 10.1016/C2012-0-00596-0.

Tankovska, H. (2021) *Leading ten accounts on Twitter in the United Kingdom (UK) as of November 2020, by number of followers*. Available at: https://www.statista.com/statistics/1009285/top-accounts-on-twitter-in-the-united-kingdom-uk/ (Accessed: 11 February 2021).

Wartmann, F. M. and Mackaness, W. A. (2020) 'Describing and mapping where people experience tranquillity. An exploration based on interviews and Flickr photographs', *Landscape Research*. doi: 10.1080/01426397.2020.1749250.

7. Biographies

Dr Phil Bartie is an assistant professor in computer science at Heriot-Watt University, where he carries out research in spatial data science. This ranges from analysing big data to gain a better understanding of space and place, to building contextually aware systems and user interfaces.

Dr Jen Dickie is a lecturer in Environmental Geography at The University of Stirling, specialising in energy geographies. She is an interdisciplinary researcher whose work focuses on understanding the complex socio-environmental interactions of the energy landscape using socio-spatial mix methods.

Dr Adam Varley is a research assistant at the University of Stirling in Biological and Environmental Science Department. He conducts research in environmental radioactivity, remote sensing and human geography and has a general enthusiasm for big data and computer programming

Dr Stacia Ryder is a Postdoctoral Researcher in the Geography Department at the University of Exeter and a co-founder of the Center for Environmental Justice (Colorado State University). Her work focuses on critical environmental justice, including her upcoming edited volume "Environmental Justice in the Anthropocene: From Unjust Presents to Just Futures."

Dr Darrick Evensen, assistant professor in environmental politics, University of Edinburgh, specialises in public perceptions and reactions to controversial energy, environment, and climate issues. He has published 50+ peer-reviewed articles in this interdisciplinary field; his research is funded by ESRC, NERC, ERC, Norwegian Research Council, and Royal Society of Edinburgh.

Patrick Devine-Wright holds a Chair in Geography at the University of Exeter and conducts research on social-psychological and spatial aspects of energy transitions. He is an IPCC lead author, Chair of the Devon Net Zero Task Force and a top 1% cited scholar for 2019 and 2020 in social science.

Professor Lorraine Whitmarsh is an environmental psychologist, specialising in perceptions and behaviour in relation to climate change, energy and transport, based at the University of Bath. She is Director of the UK Centre for Climate Change and Social Transformations (CAST), and IPCC WGII Lead Author for the Sixth Assessment Report.