

UKTwitNewsCor

UKTwitNewsCor is a comprehensive dataset, built through digital trace data on Twitter, designed to analyse local media dynamics in the UK.

Authors

Simona Bisiani¹
(s.bisiani@surrey.ac.uk), Agnes Gulyas²,
John Wihbey³,
Bahareh Heravi¹

Affiliations

¹Surrey Institute for
People-Centred AI,
University of Surrey
²Canterbury Christ
Church University
³Northeastern
University

1 Background

Local media fosters democracy and community cohesion, but decades of financial challenges have led to closures of hundreds of news outlets¹ and intense media ownership consolidation². Despite claims that this has led to **content homogenisation**, **sensationalisation**, and **reductions in local news coverage**³, there is a **lack of empirical studies of content**, resulting from **challenges in collecting data at scale**⁴ for such a broad sector.

2 Objective

To **collect comprehensive data from local news brands in the UK over a sustained period of time**, allowing researchers to:

- Analyse geographic coverage and audience engagement.
- Examine content production strategies and platform dependencies.
- Investigate the effects of external events, such as the COVID-19 pandemic, Brexit, and the cost-of-living crisis.

4 Dataset Overview

2.5M+ Articles

High Coverage

47% of all known domains (360 out of 770), 94% of LADs (337 out of 357 districts), and 39% of publishers (due to highly skewed distribution: most publishers only have one outlet).

Content Syndication Info

Duplicate and similar article cluster ids to denote groups of syndicated articles.

Geographic Metadata

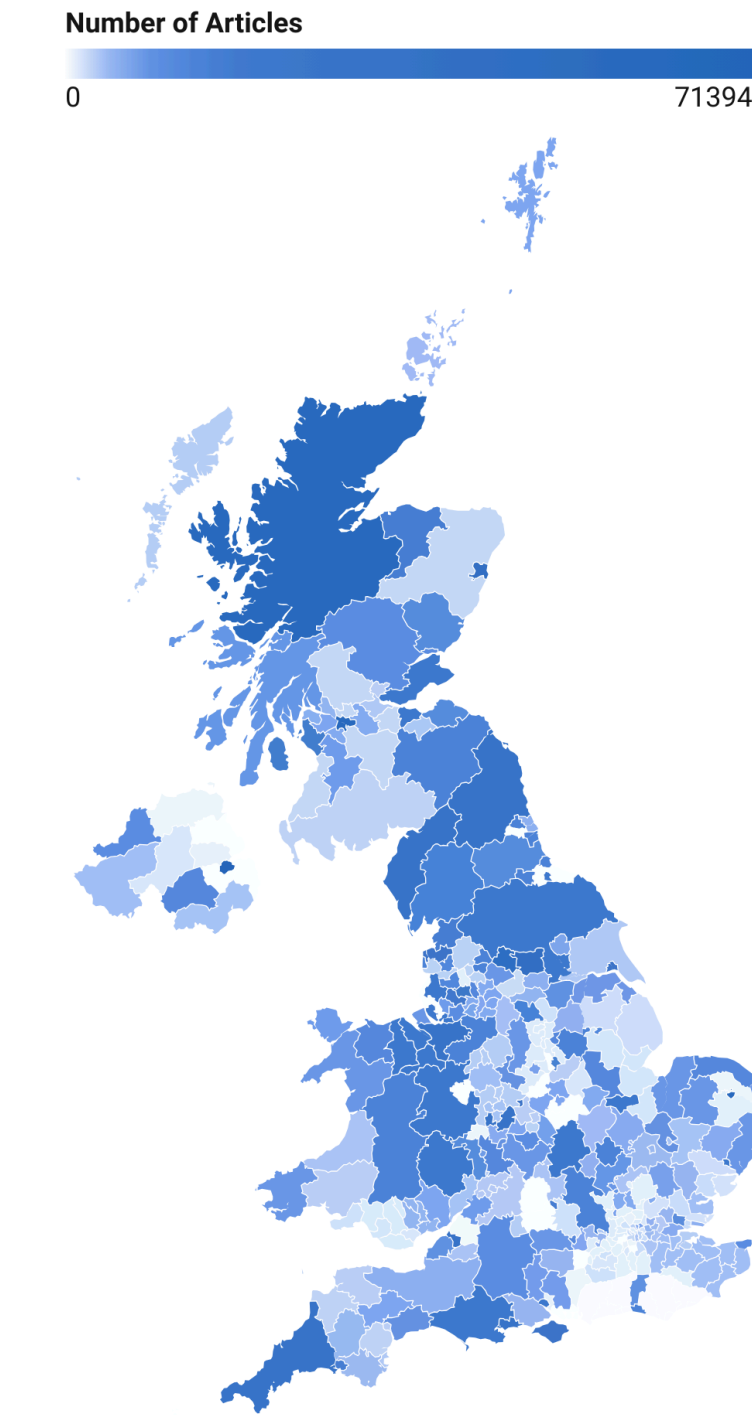
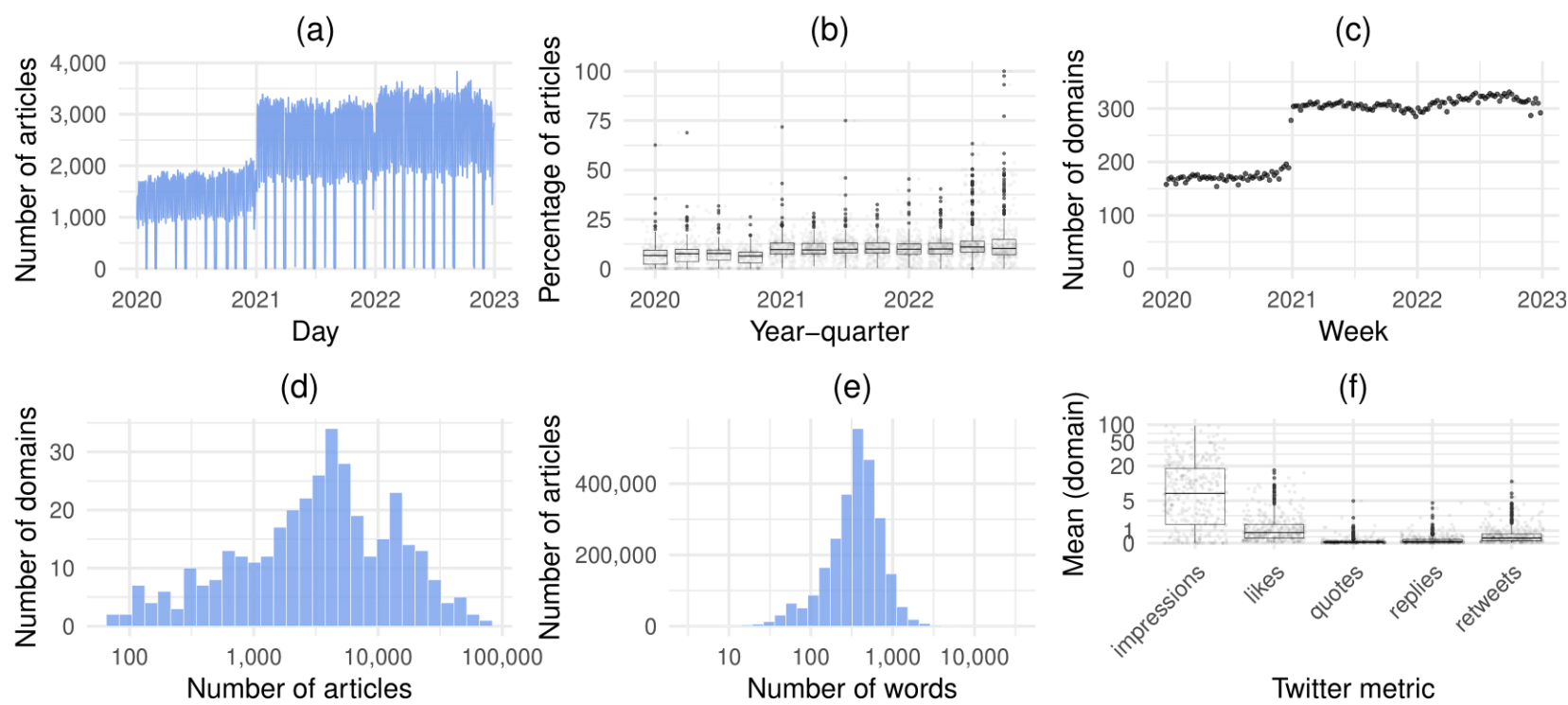
Local authority districts (LADs) of coverage for each domain.

Social Media Metrics

Tweet-level engagement metrics such as shares, likes, and comments for each article.

36 Months

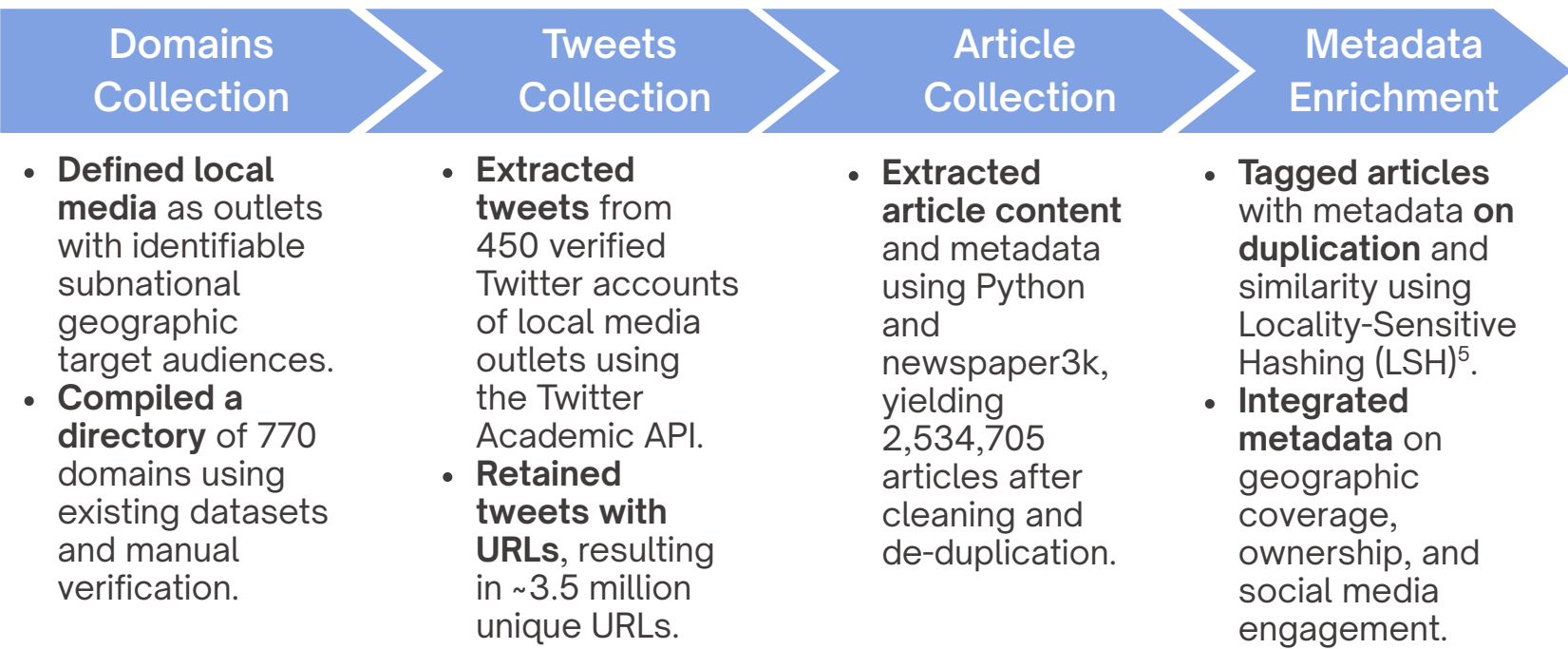
Continuous coverage from 2020 to 2022.



5 Use Cases

- Empirically define local media** by analysing the geographic scope of news coverage.
- Explore the **extent of content repurposing** within the local media ecosystem using the dataset.
- Examination of **how ownership consolidation affects** the relevance of **local news** topics. This analysis can reveal biases or shifts in reporting priorities due to media ownership structures.
- Evaluations of local news provision across the UK. It enables researchers to **identify underserved communities** and assess information provision over time.
- The dataset provides tweet-level performance metrics to study news dissemination on social media. **Factors influencing virality and audience resonance** can be identified.
- The longitudinal structure supports studies of significant historical events, including the Covid-19 pandemic and Brexit.

3 Methodology



6 Limitations

- The dataset may under-represent smaller, independent outlets and those without a Twitter presence.
- Future research should explore alternative data sources following the discontinuation of the Twitter Academic API.

Related Literature

- Hunter, T. 2024. Press Gazette.
- Media Reform Coalition. 2023. Media Reform.
- Franklin, B., Court, G., and Cushion, S. 2006. Local Journalism and Local Media, Routledge.
- Bisiani, S., and Heravi, B. 2023. Journalism and Media, 4(4).
- Jafari, O., Maurya, P., Nagarkar, P., Islam, K. M., and Crushev, C. 2021. arXiv.

Paper



Dataset



UKTwitNewsCor

UKTwitNewsCor is a comprehensive local news dataset, built through digital trace data on Twitter, designed to analyse local media dynamics in the UK.



Features

- over 2.5 million online news articles
- from 360 local media outlets
- published between 2020 and 2022
- social media performance metrics
- content duplication metadata

Covers

- 47% of local media domains
- 94% of Local Authority Districts

Allows researchers to

- empirically define local media
- explore content repurposing
- examine the effects of ownership consolidation
- evaluate local news provision
- study news dissemination on social media
- and coverage of significant events like the Covid-19 pandemic and Brexit.

