Cities 162 (2025) 105894

Contents lists available at ScienceDirect

Cities

journal homepage: www.elsevier.com/locate/cities

Exploring city dynamics through tweets: A framework for capturing urban activities as complex spatiotemporal patterns

Mahmud Tantoush *, Ulysses Sengupta, Liangxiu Han

Manchester Metropolitan University, All saints Campus, Manchester M15 6BR, United Kingdom

A R T I C L E I N F O

Keywords:
Twitter
Urban activities
Spatiotemporal patterns
Location-based social media
Complex Adaptive Systems
Bottom-up

A B S T R A C T

This paper presents a novel framework for analysing urban activities as spatiotemporal patterns using Location-Based Social Media (LBSM) data. The methodology integrates the spatial, temporal, and semantic dimensions of geolocated tweets to investigate cities as Complex Adaptive Systems (CAS) and their relationship with urban form. By combining spatiotemporal clustering (ST-DBSCAN) and topic modelling (LDA), the framework uncovers dynamic activity patterns shaped by top-down mechanisms and bottom-up self-organizing behaviours. A custom tool and Graphical User Interface was developed to support data exploration and experimentation, enabling the contextual analysis of activity clusters. The framework was tested in Manchester City Centre as an exploratory case study, focusing on the impact of Covid-19 lockdown measures as a significant disturbance. The results reveal how urban characteristics, urban form, and social behaviours influence activity levels and patterns, demonstrating fluctuations that highlight different degrees of adaptability. By exploring cities as hybrid urban-digital spaces, this approach provides an alternative approach for understanding cities as CAS, linking space to place and for exploring adaptive behaviour. The paper concludes by reflecting on the framework, use of LBSM for researching cities, and outlining directions for future work of comparing cities and integrating alternative data.

1. Introduction

Cities are Complex Adaptive Systems (CAS) that emerge from a multiplicity of interactions between actors with and within the built environment. The presence of certain land-uses and urban conditions facilitate certain types of interactions, which evolve over time. This is what Alexander et al., 1977 refers to as types of urban spaces supporting certain types of actions, and in turn actions give rise to different activities. As such, human activities and urban form of the city are highly interrelated. Local types of business, neighbourhood quality, platial and community relationships also impact how people use cities (Jacobs, 1961). Research into this interrelation, also framed as urban form and function in recent literature, seeks to understand how people interact with and within cities (Crooks et al., 2015, 2016; Jenkins et al., 2016; Smith & Crooks, 2010; Stefanidis et al., 2013). Urban form in this context typically refers to the shape of the city as a physical artifact, while urban function is more broadly defined as it relates to different definitions of social interactions and use.

Urban function is typically defined in a broader sense, whereas urban activities relate to more specific human behaviours. For example, Yuan et al. (2015) refer to urban functions as classifications of uses that meet people's different activity needs. The classifications inferred from their research include residential, historical interests/parks and commercial/entertainment areas. Crooks et al. (2015) define urban functions as the aggregation of activities that occur in urban spaces. Ye et al. (2021) follow a similar definition of functions as recurring patterns of activity over time. From these definitions, urban activities are defined as more specific types of spatiotemporal social interactions, rather than static classifications such as land use. Urban activities have also been shown to be closely related to the collective sense of place when explored through new data sources (Jenkins et al., 2016). Despite recent advancements in the literature, approaches to capture urban activity patterns at high spatial and temporal resolutions is still lacking (Niu & Silva, 2020). In this paper, place is analysed through spatiotemporal urban activities that occur within urban form as space. To better understand how urban activities and urban form are interrelated, this paper introduces and demonstrates a novel framework using Location Big Data (LocBigData) and clustering machine learning methods.

High-resolution LocBigData generated by human interactions enable research into urban activities. LocBigData refers to Big Data generated

* Corresponding author.
E-mail address: Mahmud.tantoush@mmu.ac.uk (M. Tantoush).

https://doi.org/10.1016/j.cities.2025.105894
Received 25 April 2024; Received in revised form 23 January 2025; Accepted 16 March 2025
Available online 29 March 2025

through sensors capturing human activities spatially and temporally through mobile networks, GPS, Location-Based Social Media (LBSM), location-based services, smart card travel, beacon log and camera imagery data (Huang et al., 2021). Terms such as Volunteered Geographic Information (VGI) (Goodchild, 2007), Ambient Geographic Information (AGI) (Crooks et al., 2013), Citizen Contributed Geographic (CCGI) information (Haklay, 2013) etc. are used in the literature to describe different types, applications and processes of using such data, expanding its definition from data captured through sensors to include indirect or implicit data through geo-stamps (Ballatore et al., 2024; Crooks et al., 2016). This new data environment offers the analytical possibilities to better understand the complex interactions between people with and within the built environment. Through the use of such data, as a digital "exhaust" of complex human interactions, the development of new methodologies can help provide alternative understandings of how cities function (Batty, 2016). This paper reviews different types of LBSM identifying tweets as a valuable data source for capturing dynamic urban activities, and tests a new framework that extracts spatiotemporal and semantic information to analyse urban activities as patterns of complex interactions over time.

Viewing the city as a CAS provides a framework for analysing intricate bottom-up interactions as patterns that adapt to external influences. Holland (1992) defines CAS as systems composed of interdependent components that adapt, exhibiting emergent aggregate behaviour in response to internal and external mechanisms. For cities, this theoretical framing aims at uncovering relationships between urban spaces, places and urban activities, and how top-down and bottom-up mechanisms influence them (Sengupta et al., 2016). Portugali (2006) argues that complexity holds the solution to bridge the gap between the "hard" quantitative and analytical space, and the "soft" place-based research from the humanities and social sciences. In the context of this paper, space represents urban form and conditions, whereas place relates to urban activity interactions. This complexity perspective helps to understand how local interactions give rise to aggregate behaviour as collective emergent patterns and adaptations over time for better urban planning (Rauws & De Roo, 2016). The Covid-19 pandemic has been shown to impact and change patterns of national population movements (Rowe et al., 2023). However, a methodology has not yet been developed to explore local fast dynamics of urban activities and adaptation in relation to such disruptions.

To conduct this research, geolocated tweets are collected, filtered, structured, and analysed using a custom-built tool as a methodological construct. This construct consists of three layers: a data source layer, a data analysis layer using bottom-up machine learning methods, and a Graphical User Interface (GUI). The construct is used to process activity patterns capturing count, size, locations and temporal fluctuations. The city of Manchester was selected as an exploratory test case to verify the methodology and demonstrate an alternative approach to analysing city dynamics. The analytical outcomes are demonstrated as urban activity patterns correlated with the Covid-19 lockdown measures and bottom-up changes in social behaviour. By presenting and testing this new framework, further discussions on using LocBigData to analyse, plan and design cities can be advanced.

The sections hereafter are organised as follows: The second section reviews work related to LBSM data and associated methodologies. The third section presents the novel methodology and research tool developed enabling the exploration of urban activity patterns. The fourth section introduces Manchester as an exploratory test case, and presents results generated through the tool. The fifth section discusses the use of LBSM for analysing cities as CAS, linking urban form and activities and future work for improving and applying the methodology. The sixth section summarises and concludes the paper.

2. Location-based social media and urban activities

Two main methodological considerations in using LBSM for this paper are included in this review. The first subsection reviews LBSM data sources based on an understanding of their generating mechanics and research use. The second subsection reviews methods for analysing urban activities.

2.1. Review of LBSM and their research use

Digital technologies such as GPS enabled mobile phones are becoming part of everyday life, mediating social interactions and generating digital traces as outputs (Ash et al., 2018; Kitchin & Dodge, 2011). Due to this shift, digital technologies, and data generated from daily activities and interactions, have become intertwined merging digital and urban space into a relational and multidimensional hybrid spaces (de Souza e Silva & Sheller, 2014). While urban and digital spaces were seen as separate previously (De Roo & Yamu, 2017), hybrid space present more complex relations blurring boundaries (de Souza e Silva, 2006; McLean, 2020). Analysing cities through LBSM, a better understanding of how people interact daily and engage in place-making becomes possible (Halegoua & Polson, 2021). Therefore, studying cities as physical spaces alone is insufficient. Instead, understanding cities as places that are socially co-produced and change over time through digital and urban interactions is needed (Halegoua, 2020). Cities as hybrid spaces provide an alternative approach to analysing urban spaces as reproduced, experienced and enacted (de Souza e Silva, 2009; Sengupta et al., 2020). The digital traces of complex interactions within cities as fast dynamics are captured through various LBSM digital platforms.

LBSM as digital traces can be used to understand complex social interactions, enabling possibilities for exploring how cities as CAS function. Different sources of LBSM capture different dynamics and therefore can be used to explore different research questions. Reviewing the generative mechanics of different location-based platforms enables a better understanding of how these data sources can be used to study different urban behaviours (Sengupta et al., 2020). The review of different LBSM platforms involves relating the platform mechanics to their research utility. This generates a better understanding of the types of LBSM and research purposes (see Table 1).

Different LBSM digital traces are generated through different behaviours and mechanics, producing distinct spatiotemporal traces containing different types of data. Image-based LBSM such as Instagram and Flickr tend to be useful in understanding how cities are perceived through cameras to study areas of attractions/interest (Hu et al., 2015; Kuo et al., 2018; Martí et al., 2019). The content of these images is usually directed at physical objects and locations which make it difficult to extract activity information from photos and their metadata. The photo sharing mechanics of these platforms have been used to research and explore how people are attracted to different urban areas, even in the absence of precise geo-coordinates (Burgos-Thorsen & Munk, 2023). Foursquare is based around geographically located data in the form of Points of Interest (POI)s corresponding to businesses, landmarks, amenities, etc. This mechanic means users can access comments, tips and ratings from other users to inform their travel behaviour. In this sense, Foursquare presents a collective mapping and reviews of city POIs that can be used to understand economic spatial structures, quality of services and activities through check-ins (Martí et al., 2019; Zhan et al., 2014). Facebook digital traces can be used to explore social network interactions through posts, discussions, and debates, offering insights into user and community perceptions, opinions through social network analysis. Facebook groups as a unique platform mechanic also enables research into communities linked to geographic locations identified through groups names (Ballatore et al., 2024; Madsen, 2023), providing insights into digital place-making and alternative understating of urban spaces and places.

Twitter as a platform is distinguished from other LBSM data sources in its research utility for analysing urban activities. Users of Twitter contribute short messages which can be geolocated, directly through

Table 1
Platform mechanics of different LBSM data and their research use.

Foursquare (Points of Interest)	Flickr (Photography photo-sharing)	Instagram (Lifestyle photo-sharing)	Facebook (Social networks and communities)	Twitter (Micro-blogging)
Sample of Research Use				
Land use analysis (Zhan et al., 2014) Urban deprivation (Quercia & Saez, 2014) Local habits and social behaviour (Martí et al., 2019)	Urban areas or regions of interest (Hu et al., 2015; Kuo et al., 2018) Patterns of use (Li et al., 2013) Flickr tags and land use (Yan et al., 2019)	Data Imaginaries (Burgos-Thorsen & Munk, 2023) Urban reading (Hochman & Manovich, 2013) User and group profiles (Boy & Uitermark, 2016; Martí et al., 2019)	Political Diversity (Madsen, 2023) National migration patterns (Rowe et al., 2023) Place-named groups as communications (Ballatore et al., 2024)	Mobility (Osorio-Arjona & García-Palomares, 2019) Demographics (Longley et al., 2015) Event detection (Crooks et al., 2013)
Main Platform Mechanics				
– POIs contributed by individuals and businesses. – Predefined categories are attributed to POIs. – Precise coordinate geolocation enabled.	– Photo sharing platform – Geolocated (tagging places and precise geolocation) posting of images. – Content typically focuses on photography hobbyists and professionals.	– Predominately image and video sharing platform. – Geolocated (tagging place) posts. – Content typically focuses on personal lifestyle captures.	– Social groups and networks interact through text pots and content sharing. – Geolocated posts and check-in to places and events – Contents generated through social exchanges using diverse content and interaction types	– Predominately text-based – Geolocated (tagging places and precise geolocation) posting of micro-blogs. – Geolocated content typically sharing of activities, thoughts, opinions, and business marketing.

coordinates or indirectly through data content and hashtags. Textual content of Tweets can be analysed through Natural Language Processing (NLP) algorithms that offer a more direct understanding of topics and activities spatially (Martí et al., 2019), as opposed to images in the case of Flickr and Instagram. Previous research has demonstrated the value of using Twitter data for analysing mobility patterns and travel to work (Osorio-Arjona & García-Palomares, 2019), investigating impacts and discourses of Covid-19 (Bisanzio et al., 2020; Huang et al., 2020; Iranmanesh & Alpar Atun, 2022), understanding emergencies and natural disasters (Martín et al., 2020), for developing indicators of happiness and poverty (Nguyen et al., 2016), geodemographics (Longley & Adnan, 2016), and public discourse over time (Lansley & Longley, 2016). The content of Twitter messages has been shown to contain rich qualitative information which can be processed into new types of analytical outcomes for analysing cities and their complex dynamics in various ways.

Geolocated tweets are a versatile data source demonstrated through different research applications. However, there are challenges when working with LBSM data generally (Huang et al., 2021), and specifically for different platforms. One of the main challenges often cited is the lack of representativeness. Twitter has been shown to be non-representative for traditional research use (Mislove et al., 2011). However, geolocated tweets offer the opportunity to gauge disaggregated citywide dynamics over short time horizons and are semantically rich (Batty, 2013). Moreover, tweets offer an opportunity to explore cities as hybrid spaces, where their subjective nature presents alternative analytical possibilities (Burgos-Thorsen & Munk, 2023). Research using LBSM data have been shown to be useful in better understating city life, which otherwise is not possible (Madsen, 2023; Madsen et al., 2022). This cannot be achieved by alternative traditional data sources, offering new opportunities beyond cross-sectional and expensive traditional invasive surveys. The opportunity that arises from using LBSM as digital traces of social interactions lies in better understanding cities through patterns, rather than causal modelling.

2.2. Methodologies for analysing urban activities

Research utilising LBSM data has been growing rapidly in recent years, with different methodologies applied to explore the relationship between urban form and urban activities. Jiang and Ren (2018) employed a Page Rank algorithm to describe the city's structure and hierarchy of centres using Twitter data. This approach can help reinforce or explore current perceptions of polycentricity, revealing new centres according to social media activity. Twitter data has also been used to understand sentiments as temporal sentiment fields (Kovacs-Gyori et al., 2018; Resch et al., 2016). Xing and Meng (2018) combined a supervised Random Forest method with an unsupervised Latent Dirichlet Allocation (LDA) topic modelling method to extract functional regions such as residential, public, industry, and commercial classifications, although used here for static classifications.

Explorations of temporal dynamics using LBSM data have also been conducted in different ways. McKenzie et al. (2015) used supervised Support Vector Machines and unsupervised LDA to understand the changes in Foursquare POI over time. Despite the lack of long-term POI data, observations on how the city changed over time through changes to the POIs were made. Hochman and Manovich (2013) visualised Instagram posts/images using their spectral data and temporal metadata by plotting over 30,000 posts in various ways revealing patterns and networks of activities, urban areas of interest, and user interactions. Urban dashboards have also emerged as a central topic to smart cities (Stehle & Kitchin, 2020), however this is typically designed for higher level observations such as time series for stakeholders. Social sensing is another approach used to understand socioeconomic environments. Liu et al. (2015) use this approach to understand spatial interactions by combining taxi, social media, mobile phone, into a discrete grid to analyse spatial patterns. Niu and Silva (2023) have used deep learning to infer Twitter age and gender demographics of users and highlighted their spatial differences using cluster cells. Urban Areas of Interest have been studied through snapshots using vague boundaries as a way to group individual POIs into polygon representations as collective perception mapped onto urban space (Hu et al., 2015). This approach based on a Density-Based Spatial Clustering of Applications with Noise (DBSCAN) method has been shown to be successful in capturing useful boundaries while ignoring outliers, but does not include time as an additional clustering dimension.

Geolocated tweets have also been used for event tracking and detection (Cheng & Wicks, 2014; Huang et al., 2018; Zhang & Eick, 2019). Such approaches have used space-time statistics, LDA topic modelling, clustering, and classification techniques such as density and contour-based methods. However, this research uses tweets to analyse and identify the occurrences of major events, rather than explore temporal and spatial patterns of interactions in relation to urban form. Moreover, event tracking and detection research focuses typically on large one-off events in cases of emergencies such as earthquakes (Crooks et al., 2013). Alternative approaches use dynamic topic modelling to track how salient topics discussed on social media are mapped on a regional scale (Yao & Wang, 2020). Neural Networks (Self-Organizing Maps) have also been used to provide semantic characteristics to tweets (Steiger et al., 2016). Jenkins et al. (2016) paper presents an approach to analyse urban places through tweets using LDA modelling. The six static classifications of recreation, business, education, entertainment, sports, and politics were used to give context to locations using the Getis-Ord Gi statistic.

Various methodologies have been used to explore different types of LBSM for different research topics. The methodological gap to be addressed in this paper lies in using geolocated tweets to understand urban activities as spatiotemporal patterns of complex behaviour over time. The activity patterns produced as new data outcomes as geometries for exploring adaptation, rhythms, spatial distributions, counts, sizes, locations and context rather than points or data aggregations. These types of outcomes can provide a useful contribution in understanding of urban activities as adaptive and complex, which are not readily available through traditional approaches. The outcomes can then be analysed in context to understand how top-down influences and disruptors such as Covid-19 restrictions impact bottom-up behaviour in cities as CAS. Current methods based on utilising LBSM data to analyse cities have not yet explored this framing.

3. Methodological development

This paper's main contribution lies in the development and validation of a novel framework to process tweets into dynamic urban activity patterns. As part of developing and testing the framework, a research construct as an interactive urban activities analysis tool was developed. By testing and experimenting with different data and methods, alternative knowledge and understanding of cities can be obtained challenging empirical notions (Madsen, 2023). This research approach, based in constructive research (Lukka, 2003), emphasises the importance of knowledge produced through the process of developing new constructs and through it, new understandings.

The construct developed can be abstracted into three layers: data source, data analysis, and Graphical User Interface (GUI) layer (see Fig. 1). The data sources layer feeds filtered and structured tweets data, and integrates two other data sources, land use polygons and 3D building data, directly into the GUI as a visual underlay. The data analysis layer embeds the two machine learning methods to cluster tweets and label them with the activity they best represent. The GUI layer renders the data and enables patterns generated to be exported. The GUI also includes parameters that can be tested by the user to evaluate different data outcomes, and functions to quey textual content of clsuters and temporal filtering. The subsections below discuss the three layers of the tool developed in more detail.

3.1. Data source layer

The data source layer uses three datasets: OpenStreetMap (OSM) 3D building data (OpenStreetMap contributors, 2019), the Greater Manchester Urban Historic Landscape Characterisation (GMUHLC) dataset for land uses (Redhead, 2012), and geolocated tweets. OSM and the GMUHLC data were selected due to their coverage and relevance to urban form and accessibility. OSM provides a three dimensional representation of the city, while GMUHLC offers rich insights into the functional and contextual characteristics. The OSM and GMUHLC datasets were accessed through download links and incorporated as an underlay into the GUI representing urban form. Geolocated tweets were collected through the free Twitter Application Programming Interface (API) service were used for capturing urban activity patterns.

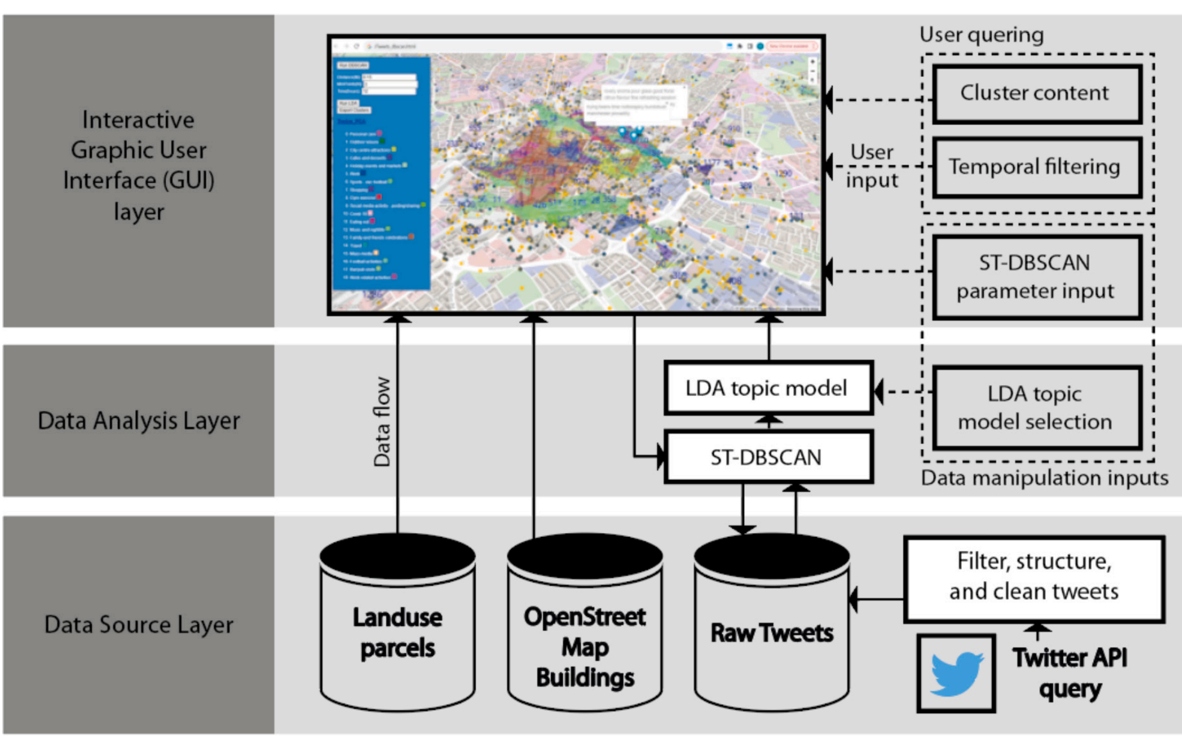

Fig. 1. Research tool developed for capturing urban activity patterns from geolocated tweets. The tool comprises of three layers: Data source, data analysis and interactive graphic user interface layer.

3.1.1. Twitter API access and representativeness

Recent changes to Twitter's API access, including recent restriction of free tier usage, have created significant challenges for researchers. These restrictions are part of a broader trend, referred to as the "API-calypse" (Bruns, 2019), in which major platforms like Twitter and Facebook have increasingly limited API access to protect user data and prevent misuse. Prior to 2019, as part of Twitter's free streaming API terms and conditions, Twitter allowed 1 % of their tweets to be downloaded through their services. Since geolocated tweets make up a similar percentage of tweets, geolocated tweets could be filtered and directly accessed through the Streaming API (Longley & Adnan, 2016). In July 2019, Twitter announced that it will be disabling the option to geolocate tweets directly through the twitter application as most users did not use the precise geotagging feature (Shu, 2019). Geotagging is still used and available through Twitter where people can tag to named POI locations linked to coordinates, but precise geolocating through the Twitter app was discontinued.

Precise geolocating features in Twitter are still available through third-party linked apps such as Instagram, Foursquare, and the Twitter camera features. Hu and Wang (2020) conducted an analysis of three large Twitter datasets and found that the discontinuing of precise geolocating through the Twitter app was not a major loss for the research community. The data sets analysed by Hu and Wang (2020) revealed that between 8 % to 25 % of precise tweets came from the Twitter app directly, whereas 92 % to 75 % came from third-party applications. While the percentage of precise geolocated tweets have been reduced, geotagging to POIs are still available and used in this research in addition to precise coordinates. A small percentage of tweets include geolocation metadata, raising concerns about representativeness. Geotagged tweets often come from younger, tech-savvy users or those engaged in public-facing activities, introducing bias (Mislove et al., 2011). While this limits broader generalisation, Twitter remains a valuable dataset for analysing city dynamics.

3.1.2. Collecting and structuring tweets

The overall process from the API query to cleaning tweets in this paper is presented below (Fig. 2). The process involves weekly API queries, which filter and manipulate tweet messages and semantic content into an analysable format. The workflow for collecting tweets includes querying the Twitter servers, removing personal information, cleaning, and manipulating the data into a machine analysable data structure. Unnecessary and personal information such as Twitter handles are removed. Data is cleaned by removing duplicates and tweets with inaccurate geolocation information. Tweets from hyperactive Twitter accounts, often associated with bots, were excluded from this research to avoid skewing the data. While bots can provide valuable insights into digital spaces and patterns of information exchange (Marres, 2015), they were considered noise in this paper. This is because bots typically generate automated content that does not reflect urban activities.

The unstructured format of tweets, as human messages that can contain sentences and emojis, require processing into an analysable

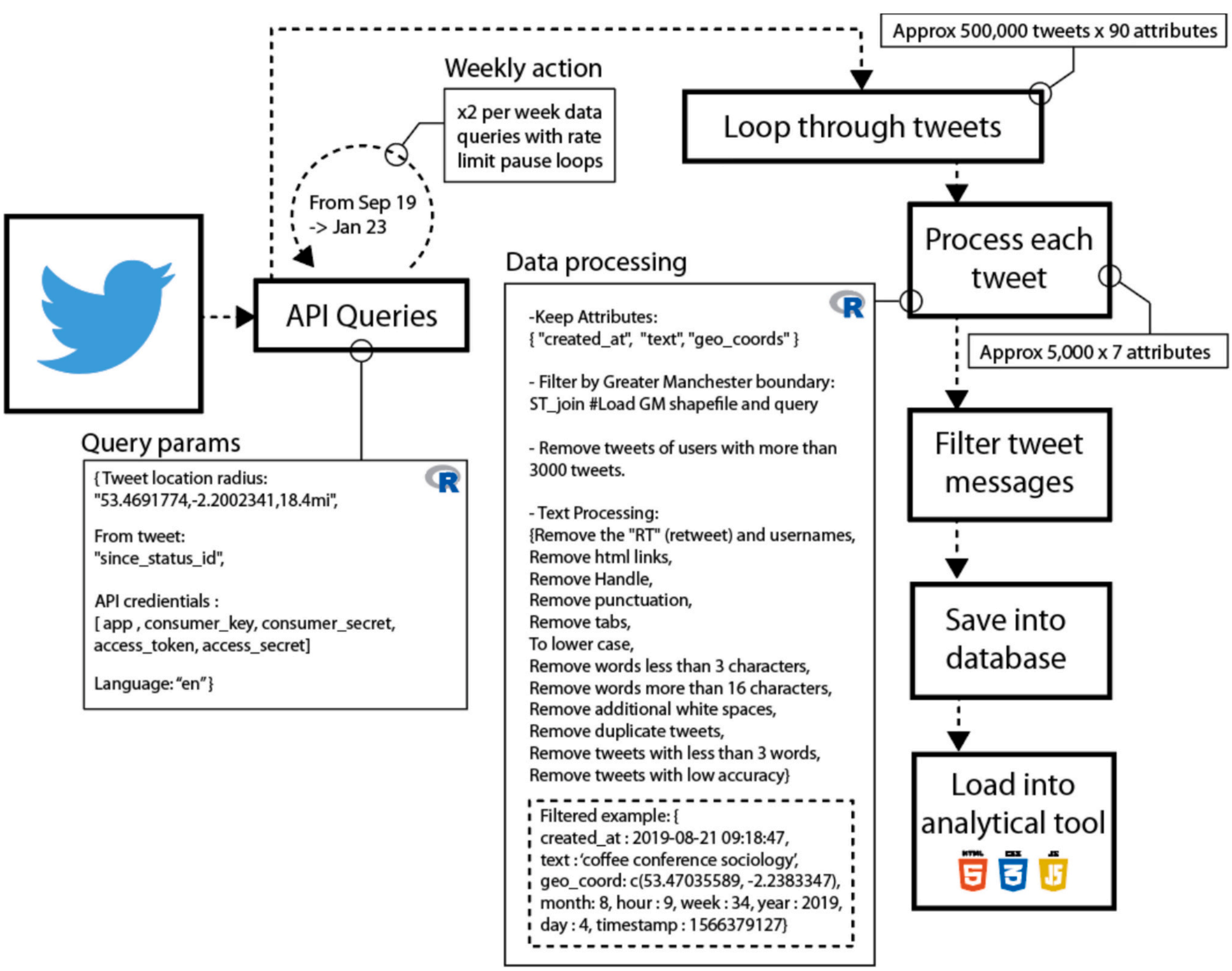

Fig. 2. Workflow of collecting tweets twice a week and processing into structured and machine analysable tokens.

format as a list of individual tokenised words as vectors. The process of transforming words to vectors (Word2Vector) creates a list of important simplified words excluding stop words such as “the”, “and” or “a”. These simplified words are processed from the original format to exclude noise and unnecessary information. Words need to be transformed into vectors representing semantic meaning, and thus the process of structuring tweets into clean machine processable data is needed. This process uses the lemmatisation toolkits developed by Manning et al. (2014), and best practice examples from the literature (Lansley & Longley, 2016; Resch et al., 2018). The following stages were followed specifically for structuring tweets:

1- **Remove retweets:** removing all retweets as they do not pertain to a specific location or activity in most cases.
2- **Remove HTML links:** links to other webpages are removed as they do not contain any analysable semantic information relevant to urban activities.
3- **Remove handles:** the handles of any other accounts included in the tweet are removed as they do not contain any activities information.
4- **Remove punctuation and hashtags:** full stops, question and exclamation marks for example are characters which do not hold any semantic information in themselves.
5- **Remove spaces and tabs:** extra spaces are removed as their inclusion as a character would identify them as a different word and would skew the analysis.
6- **To lower case:** the text is all transformed into lowercase, as words containing uppercase letters would be processed as a different word skewing the analysis.
7- **Remove words less than three characters:** these words are often overused and don’t hold useful semantic information.
8- **Remove words more than sixteen characters:** words with more than 16 letters are often mistakes or very specific words which could skew the analysis.
9- **Remove duplicate tweets:** some tweets can be duplicated and are removed to prevent skewing the topic modelling.
10- **Remove tweets less than three words long:** three word tweets contain insufficient semantic information which can skew the analysis.
11- **Remove tweets with low geographic accuracy:** removing tweets with coordinates less than four decimal places as these are not precise enough.

The API returned around 140 million tweets between September 2019 and January 2023 for a search radius containing Greater Manchester. Based on the data processing steps of filtering tweets using geolocation and text-based rules, a total sample of 141,002 tweets in Greater Manchester were obtained and used to develop the methodology and demonstrate results.

3.2. Data analysis layer

Two clustering machine learning methods are used together to extract urban activity patterns from tweets. The first method, Spatial-Temporal Density-Based Spatial Clustering of Applications with Noise (ST-DBSCAN), is used for spatiotemporal clustering. A trained LDA model is then used to label each spatiotemporal cluster with the highest weighted LDA activity topic, representing the most prominent theme within that cluster. The trained topic model used was modelled on all the tweets collected, which captured 19 different activity types unique to the data sample.

3.2.1. Spatiotemporal clustering

Density-based clustering was identified as the most suitable method for extracting urban activity clusters from tweets. This is due to the method’s ability to ignore outliers, be applied from the bottom-up without a pre-defined number of clusters, incorporate additional distance parameters such as time, and efficiency with large datasets. The DBSCAN method was originally developed by Ester et al. (1996), and used for different research applications. The uses include understanding different regions of interest using Flickr data (Kuo et al., 2018), photography clusters using Flickr (Kisilevich et al., 2010) and cognitive regions by combining datasets such as Instagram and Flickr (Gao et al., 2017). The adaption of the method used in this paper for tweets is ST-DBSCAN. This is a development of the original DBSCAN algorithm developed by Birant and Kut (2007), which incorporates time as an additional distance parameter. This works in the same manner as the Euclidean distance parameter, but distance is measured as seconds rather than meters. In this technique, EPS1 is the minimum distance in meters between points, and EPS2 is the minimum distance in seconds between points (see Fig. 3).

3.2.2. LDA topic modelling for semantic analysis

LDA is considered the most popular topic modelling approach first introduced by Blei et al. (2003). LDA topic models are used to extract and understand semantic text data, relationship between a document and text, and to use computers to structure and add meaning to documents (Jelodar et al., 2019). While sentiment analysis enables the research to extract positive and negative emotions from text (Gandomi & Haider, 2015), topic modelling is used to extract topics, or underlying themes, from a collection of texts to add structure and infer meaning. LDA topic modelling has been used extensively to analyse geolocated text to understand geo topics (Tenney et al., 2019; Yao & Wang, 2020), spatiotemporal event detection (Cheng et al., 2016; Zhang & Eick, 2019), identify land uses (Frias-Martinez et al., 2012), Geodemographics (Longley et al., 2015; Singleton & Spielman, 2014), among other applications. LDA is a generative probabilistic model used in this paper as an unsupervised approach to extract urban activity topics from tweets. While, topics can be analysed and extracted by a researcher over a small collection of documents, analysing thousands of text documents is only possible using an algorithmic approach to summarise topics and relationships as multidimensional vectors (Chauhan & Shah, 2022). The use of LDA topic modelling for tweets in this paper is demonstrated below (see Fig. 4), mapped against the LDA model parameters (Blei et al., 2003).

Different model iterations using the tweets collected were tested. A range between 16 and 37 were topics was evaluated using intrinsic coherence scores (Newman et al., 2010), and extrinsic human judgment (natural cluster evaluation) using the PyLDAvis tool (Sievert and Shirley, 2015) to examine each model (see Fig. 5). Intrinsic and extrinsic measures are part of the topic modelling process and used to inform model selection and improvement (Chang et al., 2009). The human judgment process involved using visualisation technique such as word clouds and topic distributions (Sievert & Shirley, 2015) to examine word and topic intrusions (Chang et al., 2009). The extrinsic approach investigates topics within a range of models to identify the most suitable model for natural human clusters. The 16-topic model captured broad activities and lacked granularity, while the 37-topic model overfit the data identifying specific events. A model consisting of nineteen topics was selected as the most suitable model for the tweet sample collected based on iterative evaluation by exploring the topic distribution for each model. Each topic is then given a label as an activity name it best represents based on the activity theme captured, such as “Eating Out”, “Attractions” or “Football Activities”.

Since LDA is probabilistic, each ST-DBSCAN cluster was assigned a distribution comprising the 19 topics representing urban activities, with the most prominent topic for each cluster designated as the main label for ease of interpretation. The full probability distribution, however, was retained to inform the researchers through the analysis stage. The probability distribution is not a measure of predictive accuracy as the LDA model accounts for linguistic variation, as terms can occur across multiple topics. For example, the word “drink” may have high

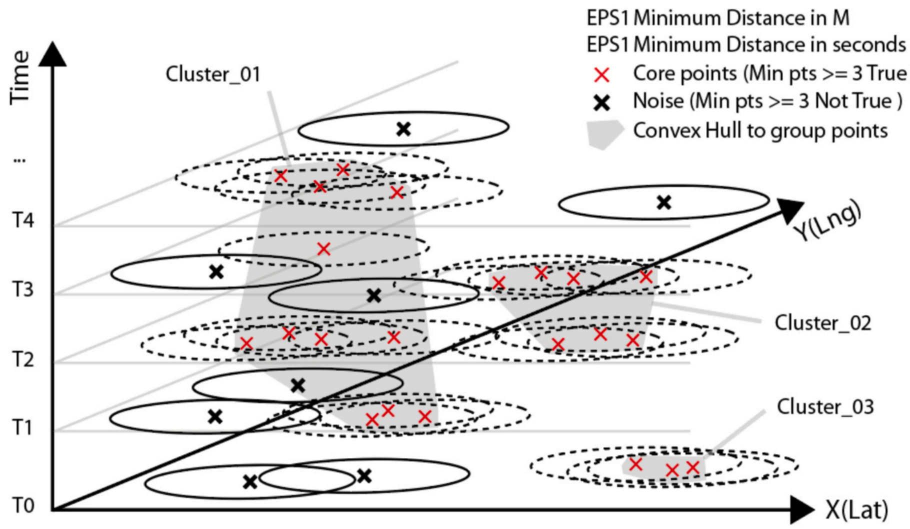

Fig. 3. Spatiotemporal clustering using the ST-DBSCAN technique, including three dimensions: latitude (X-axis), longitude (Y-axis), and time (vertical axis). The figure demonstrates how ST-DBSCAN identifies spatiotemporal clusters. Adapted from Birant and Kut (2007).

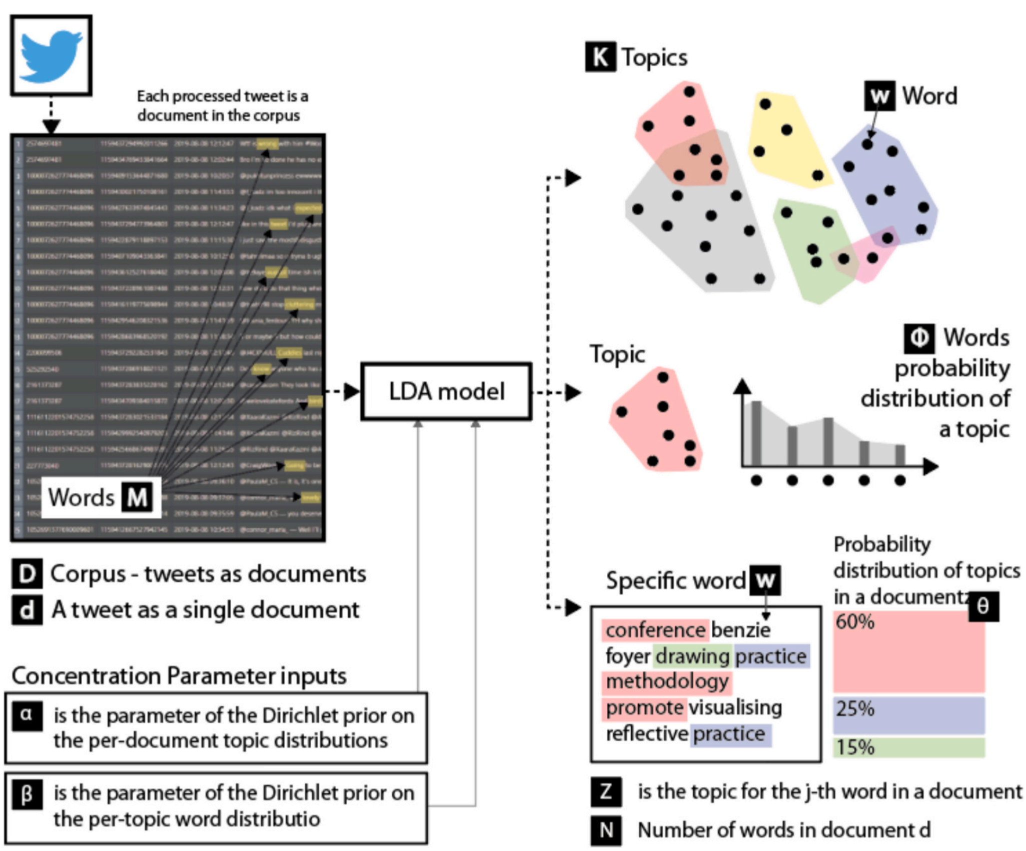

Fig. 4. A representation of the LDA modelling process applied to a corpus of tweets, illustrating key components and parameters. The dataset (D) where each tweet is treated as a single document (d), forming the collection of documents. Topics (K) are identified from tweets as types of urban activities.

probabilities for both the “Cafes and Desserts” and “Pub Visits” topics, highlighting overlapping meanings rather than a definitive association with a single activity. These considerations enable nuanced interpretations of urban activity patterns while recognising the inherent complexities of language use.

3.3. Graphic user interface layer

The GUI was designed to allow researchers to interact and experiment with urban activity patterns by testing parameters and interrogating their semantic content (see Fig. 6). This interface design enables testing of different parameters to explore different rhythms as data experiments (Madsen, 2023), allowing for alternative patterns to be generated. The spatiotemporal cluster patterns generated using the ST-

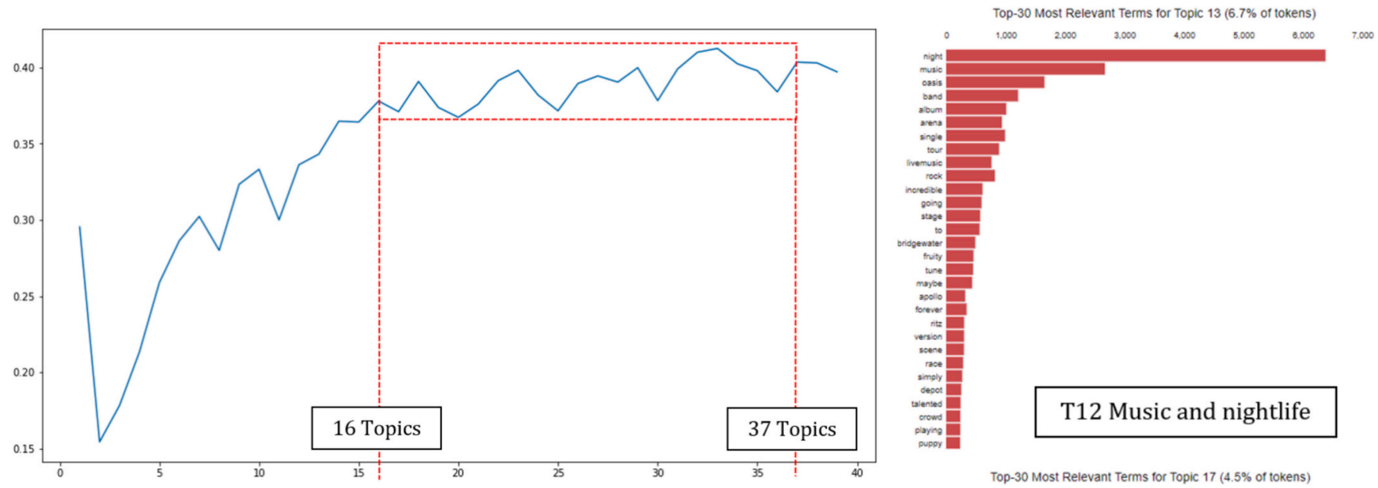

Fig. 5. Identifying range of relevant topics using coherence values. X axis refers to the number of topics in the model, and Y axis to the coherence value (left). Using the PyLDAvis tool for human judgment to assess cluster contents (right).

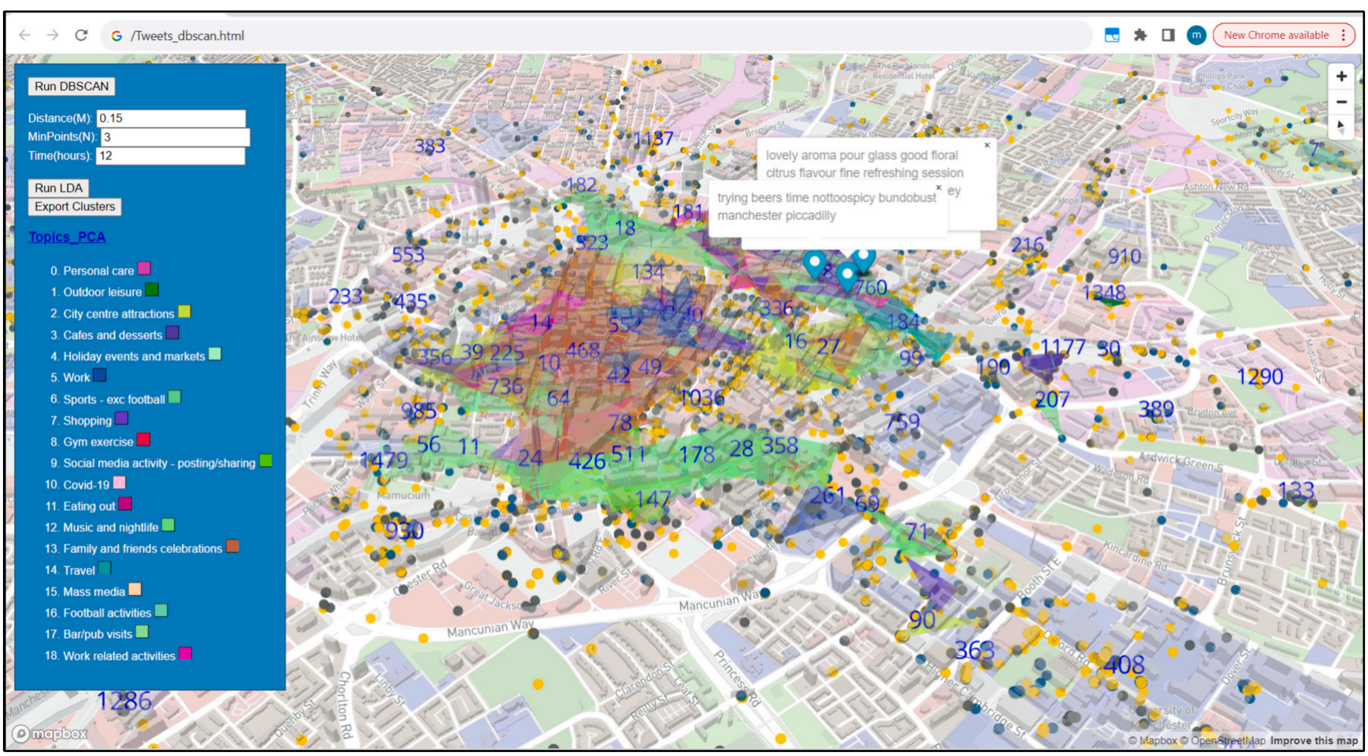

Fig. 6. A bespoke rtesearch tool developed to process spatiotemporal activity patterns from geolocated tweets. The screenshot showcases how ST-DBSCAN and a LDA topic model are combined within the user interface. The extracted topics from tweet clusters are categorised into 19 different activities. Each topic is color-coded using the integrated legend for visualisation.

DBSCAN method are visualised and labelled with an activity for ease of interpretation. Data exploration and interrogation are facilitated by filters and popups, and a function to export patterns generated in a structured JSON format is also provided.

The theoretical approach of this framework is grounded in pattern language theory (Alexander et al., 1977) and cities as CAS (Holland, 1992). Both combined enable the exploration of cities as emergent patterns that adapt overtime, linking space as urban form to places as urban activities representing complex social interactions within cities (Portugali, 2006). A patterns based approach enables the exploration of geometries, providing unique analytical possibilities on the relationship between form and function using LocBigData (Crooks et al., 2015; Huang et al., 2021). A patterns approach is distinct in its ability to present different spatial and temporal characteristics, such as rhythms, sizes, counts, distributions and frequencies, which can be tracked over time to reveal city dynamics.

4. Capturing city dynamics through spatiotemporal activity patterns

This section presents the results of demonstrating the methodology through four subsections. The first introduces Manchester City Centre as an exploratory test case. The second explores the effect of clustering parameter selection on the generated patterns. The third examines temporal dynamics captured in relation to Covid-19 lockdown measures. The final subsection discusses the spatial distribution of activity patterns against urban form.

4.1. Manchester city centre as exploratory case study

Manchester was chosen as an exploratory test case to validate the proposed methodology. As an exploratory case study, the choice of Manchester serves to demonstrate the framework's applicability and identify avenues for future research (Yin, 2014).

A map of central Manchester is provided to contextualise the analysis (see Fig. 7). Manchester City Centre is the regional centre of Greater Manchester. The Oxford Road corridor is considered a spine of the city linking large institutions (City Centre to Manchester Metropolitan University, University of Manchester, and the main regional hospital). The area is annotated according to the functional characteristics of different districts/quarters. Piccadilly Gardens, St Peter's Square, and the Arndale Shopping Centre are considered focal points. The shopping district (North) hosts most of the high street shops across two axes from Piccadilly Gardens and towards Deansgate. Manchester City Centre presents a dense and mixed-use urban form with functions ranging from commercial offices (A), entertainment (B), bars and restaurants (C), and shopping (D).

The Northern Quarter (East) is considered a lively and popular area with many independent clothing shops, bars, music venues and cafes. China Town is situated in the centre with many ethnic restaurants and international shops. The Gay Village is south of China Town, with numerous bars and a popular night scene. The conference centre is where the Manchester Central Convention Complex is located along with hotels. The cultural quarter (West) hosts many of the city's popular museums, including The Science and Industry Museum, and restaurants. Deansgate is a popular area situated between the shopping and cultural areas, known for restaurants and presence of large commercial office buildings. The urban form of Manchester City Centre presents a mixed and diverse range of buildings and urban typologies supporting a wide range of activities.

4.2. Parameter selection and testing of pattern outcomes

The ST-DBSCAN clustering algorithm was used to generate activity clusters across Greater Manchester, with parameters of EPS1 = 0.15 km, EPS2 = 12 h, and MinPts = 5 yielding over 4000 activity clusters. Of these, 1666 clusters occurred in Manchester City Centre. These parameters were iteratively tested to balance cluster granularity and scale. An EPS1 parameter less than 0.15 km was found to produce small and specific clusters, whereas a larger distance value produced more general large clusters which are not of appropriate scale. EPS2 set to 12 h effectively captured single day activities, and a minimum of five points per cluster reduced noise, resulting in meaningful activity pattern outcomes (see Fig. 8). These parameters were selected to demonstrate the methodology for capturing urban activity patterns. However, alternative parameters and data could be explored to capture different temporal rhythms and spatial scales.

From the geolocated tweets, 19 distinct topics representing activity types were extracted (see Fig. 9). Manchester City Centre's functional capacity in relation to the region is highlighted. For example, the "Music and Nightlife" activity was most prominent in the City Centre capturing 58 % of the total activity count in the region. Other dominant activities

Fig. 7. Manchester City Centre context with key areas annotated. This map highlights various zones of interest, marked with specific annotations (A-D) that correspond to the street view images for visual context. Landmarks such as the high street shopping areas, Piccadilly Gardens, and Oxford Road Corridor are prominently labelled.

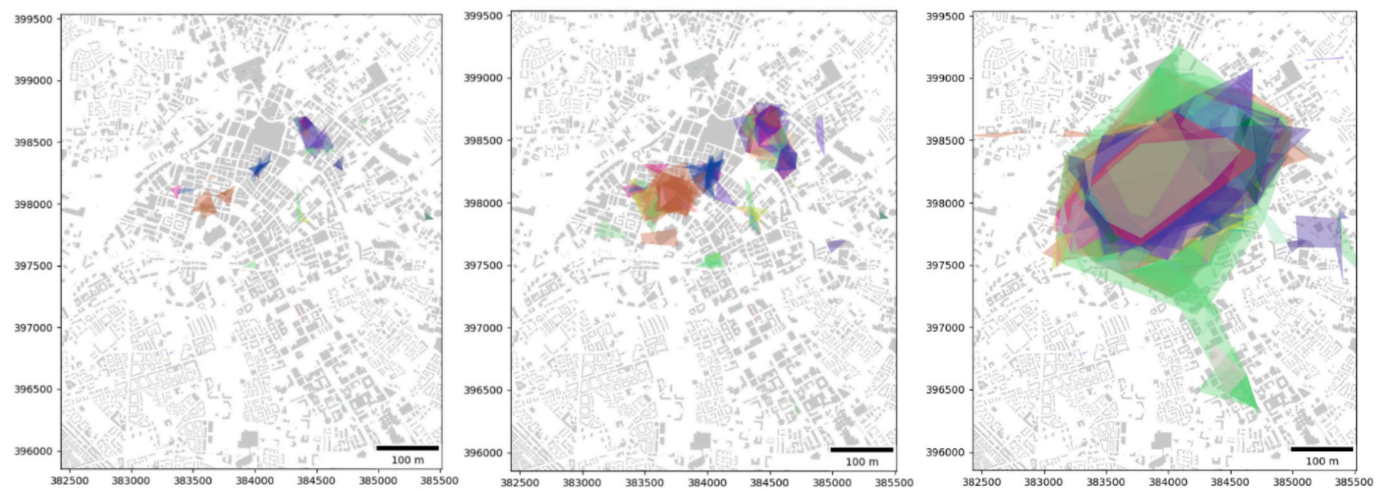

Fig. 8. A visualisation of spatiotemporal clustering results using different EPS1 parameters for the ST-DBSCAN algorithm: 0.05 km (left), 0.15 km (centre) and 0.25 km (right).

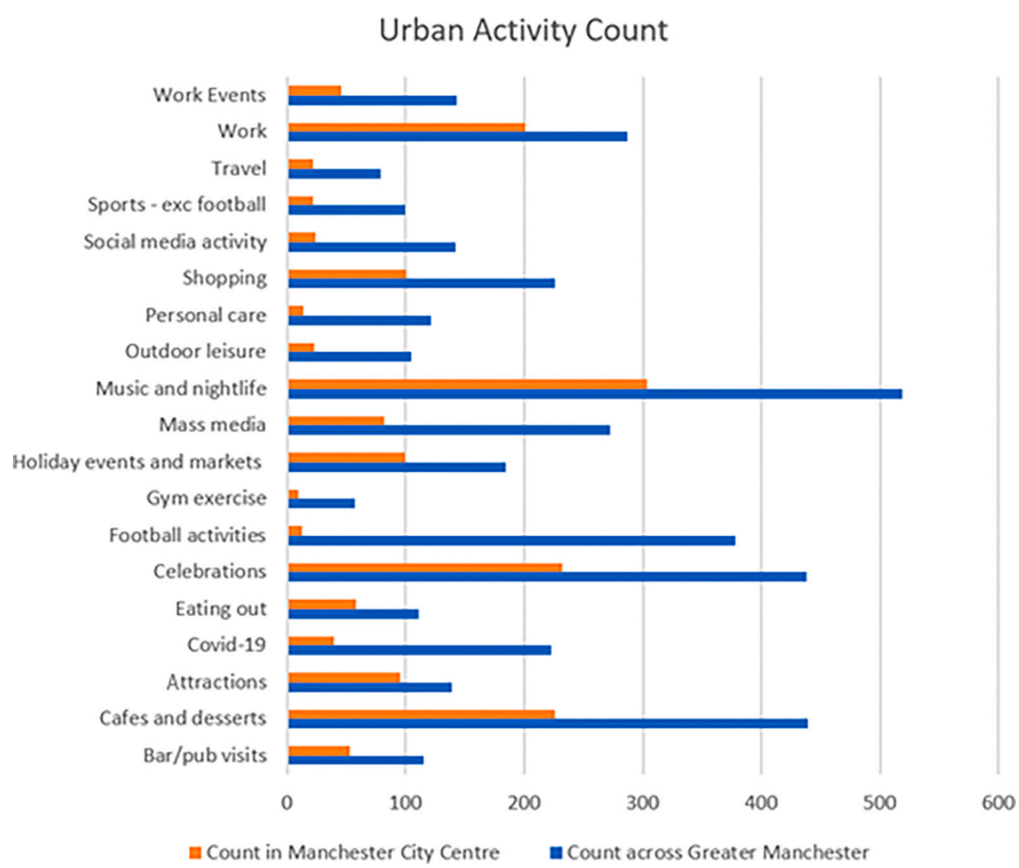

Fig. 9. Topics extracted as bottom-up urban activity types and their count in Greater Manchester compared to Manchester City Centre.

included “Celebrations”, “Cafes and Desserts”, and “Work” all heavily concentrated in the City Centre. Conversely, activities such as “Football Activities” and “Travel” were less represented in the City Centre, aligning with their association to stadiums and transport infrastructure outside the centre.

Urban activity clusters are prominent in the dense urban centre due to higher tweet densities. As the regional hub, Manchester City Centre is most active area, with approximately 40 % of all activities occurring within its boundaries. While the bar chart representation limits the analysis to counts, the results highlight Manchester City Centre as the focal point for further investigation. Focusing on this area as an exploratory test case allows for more granular analysis into the patterns generated and their dynamics.

4.3. Temporalities of urban activity patterns in Manchester City Centre and impact of Covid-19 lockdown measures

The analysis of activity patterns in Manchester over time illustrates the interplay between urban spaces and places through a complexity lens using digital traces. This theoretical lens helps to interpret the relationship between urban form, place, and activity, and how top-down regulations interact with bottom-up social behaviours (Portugali, 2006; Sengupta et al., 2016). The Covid-19 pandemic provides unique conditions for understanding these dynamics, as national and local lockdown restrictions significantly altered patterns of movement and social interactions at the time (Rowe et al., 2023).

Urban spaces inherently facilitate or restrict activities based on their form and quality (Alexander, 1979; Jacobs, 1961), but an understanding of how city dynamics adapt is still lacking. This became evident during the Covid-19 pandemic, when lockdown measures altered regular activity patterns. To contextualise the analysis, the timeline of restrictions provided by the Institute for Government (2022) was used. The initial lockdown on 23rd March 2020 led to a steep decline in activity patterns captured, followed by fluctuations as restrictions were adjusted. Key milestones, such as the easing of restrictions in summer 2020 and the phased “Roadmap Out of Lockdown” in 2021, marked shifts in activity patterns. The results revealed how the different phases of lockdown correlated with activity fluctuations in Manchester City Centre (see Fig. 10). Before restrictions, November 2019 recorded over 175 activity clusters, with “Celebrations” and “Music and Nightlife” dominating the urban landscape. These clusters dropped sharply to around 25 per month during the first lockdown, underscoring the immediate impact of social distancing measures. As restrictions eased, certain activities, such as “Cafes and Desserts”, exhibited notable peaks, illustrating their sensitivity to top-down policy changes. Conversely, activities like “Work” gained prominence after restrictions were lifted, potentially reflecting public discourse on returning to workplaces.

Patterns of fluctuation varied significantly by activity type. Social activities tied to specific locations, such as “Bar/Pub Visits”, “Eating Out”, and “Music and Nightlife”, experienced severe declines during lockdowns, with slow, incremental recovery during phased reopening. In contrast, “Mass Media” and “Social Media Activity” remained stable or increased, reflecting a shift towards digital engagement as individuals sought alternative ways to connect and stay informed. Outdoor activities like “Outdoor Leisure” recovered moderately during easing phases, indicating a preference for safer, outdoor environments. Other activities, such as “Football Activities” and “Shopping”, exhibited partial recovery tied to specific reopening milestones, while “Travel” and “Holiday Events and Markets” faced prolonged disruptions due to mobility restrictions. Incremental recovery was observed in activities like “Celebrations”, reflecting the gradual social and economic adjustments made under changing regulations.

These results illustrate an example of the adaptive capacity cities as CAS, where the dynamic interplay between top-down regulations and bottom-up behaviours shaped recovery trajectories. Activities such as “Music and Nightlife” and “Cafes and Desserts” demonstrated resilience, with their prominence fluctuating alongside restrictions, while others maintained stability or showed delayed recovery. The diverse rates of recovery across activities underscore the potential of Location-Based Social Media data to uncover nuanced patterns of adaptation. The framework demonstrates a valuable opportunity to explore adaptive patterns for generating insights into how cities function as CAS.

4.4. Spatial urban activity patterns and urban form

The spatial mapping of urban activities reveals distinct relationships between activity patterns and urban forms, aligning with Alexander et al.’s pattern theory (Alexander, 1979; Alexander et al., 1977) of urban spaces supporting specific actions, which in turn give rise to distinct activities. This interrelation highlights how local businesses, neighbourhood quality, and spatial relationships influence urban vitality (Jacobs, 1961). Urban activity patterns, when captured as spatiotemporal clusters, provide an alternative way to understand and analyse cities. By identifying which activities dominate specific areas and how they vary over time, the methodology and patterns generated can uncover emergent behaviours that reflect the interplay between urban form and function (Crooks et al., 2015).

The spatial patterns captured (see Fig. 11) demonstrated that

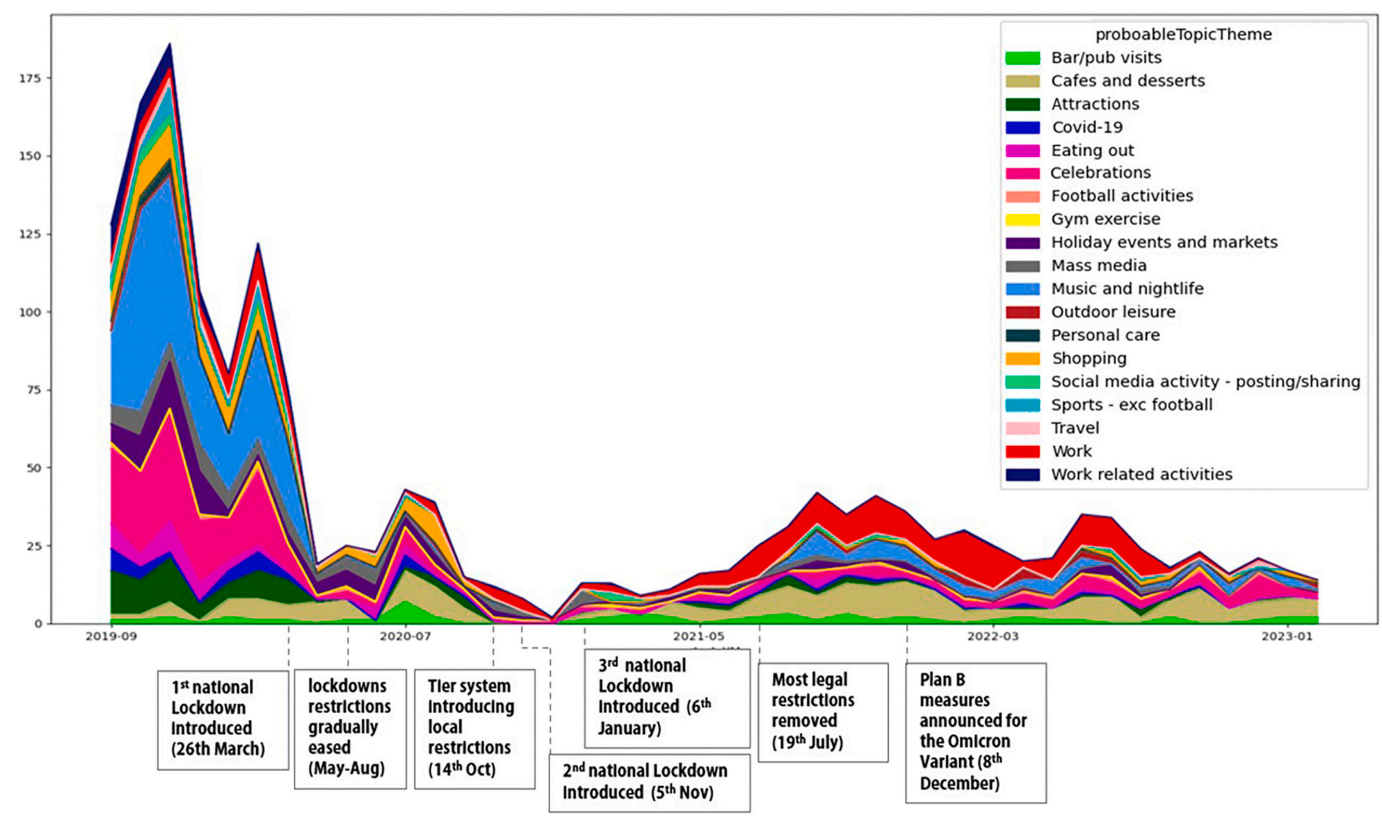

Fig. 10. Urban activity patterns in Manchester City Centre, grouped by month to reveal temporal adaptation to national and local lockdowns.

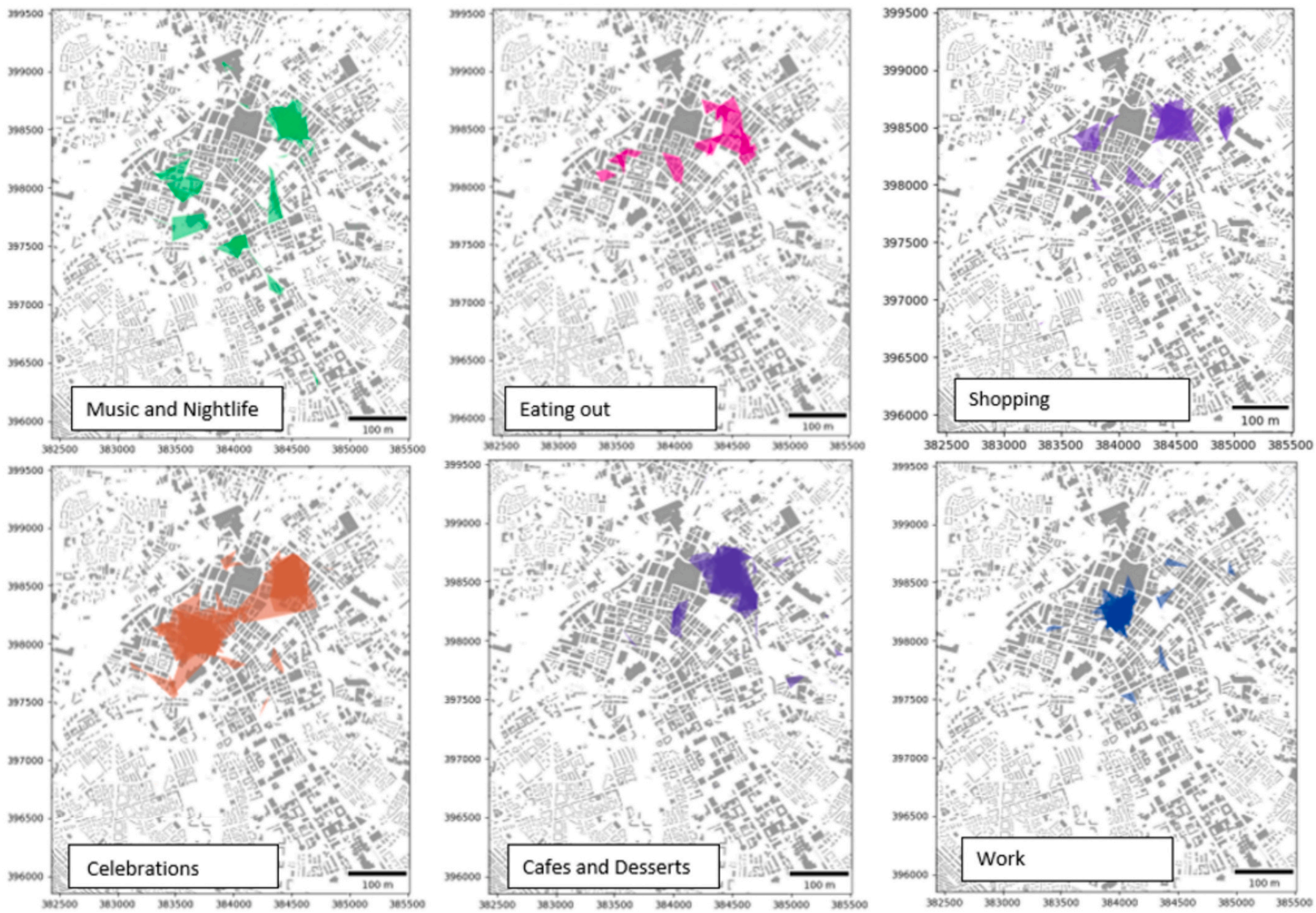

Fig. 11. Mapping prominent urban activities as spatial patterns in relation to urban form in Manchester City Centre.

activities like "Eating Out" formed larger, concentrated clusters, with hotspots in Deansgate, Chinatown, and the Northern Quarter. In contrast, "Music and Nightlife" activities were more dispersed, appearing around key venues such as the Northern Quarter, Manchester Arena, and Deansgate. "Celebrations" were prominent throughout the area, particularly in commercial areas, whereas "Work" activities were unexpectedly centralised. Interestingly, areas dominated by independent shops, cafes, and restaurants, such as Deansgate and the Northern Quarter, exhibited more vibrant and complex activity patterns compared to high street shopping areas like the Arndale Shopping Centre, raising questions related to places and digital place making (Halegoua & Polson, 2021). By exploring the spatial characteristics of different urban activities, an alternative approach is demonstrated that combines 'hard' spaces (urban form) and 'soft' places (activity patterns) through a complexity lens as suggested by Portugali (2006).

The Northern Quarter emerged as a key focal point for urban activity, reflecting its diverse and creative role within the city. The results suggest stronger sentiments towards this area expressed on Twitter, indicating its unique sense of place with positive social interactions such as "Celebrations" and "Music and Nightlife" activities. Tweets referenced its independent businesses, cultural vibrancy, and distinctive character, which contrast sharply with the limited activity observed in the Arndale Shopping Centre, despite its size and significance as a regional commercial area. This raises questions about the role of independent businesses and cafes in fostering urban vitality, aligning with Jacobs' (1961) emphasis on the importance of small-scale, diverse enterprises. Such granular insights into platial relationships offer new opportunities to explore the social dynamics of urban spaces (Jenkins et al., 2016). The pattern generated illustrate how different urban forms facilitate specific types of activities and foster varying degrees of engagement. Areas with a mix of independent shops and public spaces support richer hybrid space interactions, where urban and digital behaviours converge (de Souza e Silva, 2006). For instance, the Northern Quarter's prominence reflects its ability to attract and sustain diverse activities, creating a dynamic and adaptive environment as was captured through tweets. In contrast, more homogenised spaces like the Arndale Shopping Centre exhibit limited activity diversity and engagement.

Further exploration into the relationship between urban conditions, sense of place, and vitality could refine these exploratory findings. By examining activity size, shape, and location over time, the methodology and tool presented enable an alternative understanding of city dynamics as patterns of activities facilitated by different urban forms and conditions. This framework can also help address challenges in LocBigData research by advancing alternative analytical possibilities (Huang et al., 2021). The results presented demonstrate the methodology and highlights its contributions to current approaches on analysing urban form and function, future research, and potential applications for planners, designers, policymakers.

5. Discussion

The framework and methodology presented were tested and explored through the spatiotemporal urban activity patterns generated. This section revisits the discussions related to LBSM data use, the theoretical underpinnings of linking urban form and function through a complexity approach, broader applicability of the framework and future methodological considerations.

5.1. Use of LBSM for exploring city dynamics

This paper demonstrated an alternative approach of using LBSM, particularly geolocated tweets, for analysing city dynamics by linking urban form as space and urban activities as place. Geolocated tweets offer an alternative lens to study cities as complex urban patterns, though biases related to user demographics must be critically considered (Mislove et al., 2011). For instance, Twitter's predominantly younger user base may disproportionately reflect activities like "Music and Nightlife", whereas a different demographic composition might reveal a broader or alternative activities. Future research could integrate demographic data with urban activity analysis to provide a more holistic understanding of engagement patterns and explore how various population groups interact with urban spaces differently.

Geolocated tweets and their content facilitated spatiotemporal and semantic analysis, enabling insights into dynamic activity patterns. Alternative methods for extracting and combining Twitter data with other LBSM sources will provide alternative insights. Each LBSM source is generated through unique platform mechanics that can be combined and applied to different research questions. For instance, a patterns approach can be applied to social networks and group dynamics using Facebook data, Foursquare's predefined POI data is effective for capturing patterns of landuse, and Instagram's or Flickr's rich photo lifestyle focus can be used for visual patterns. Combining these with Twitter's textual richness can provide new insights into hybrid urban-digital spaces.

The results suggest that digital spaces and place making processes are intricately tied to urban form. Digital traces, such as tweets, not only reflect activities but also shape and reinforce perceptions of urban spaces (Halegoua, 2020; Halegoua & Polson, 2021). The merging of urban and digital spaces influences how people perceive and experience cities (Sengupta et al., 2020). For instance, the prominence of "Music and Nightlife" activities in Manchester reflects the city's cultural identity but prompts questions about its uniqueness compared to other cities. Future methodological development and research could refine the methodology to facilitate comparative analysis, identifying distinct urban activity profiles for different cities.

5.2. Linking urban form and urban activities for adpative policy and design

The framework demonstrated that different urban forms and their characteristics are highly interrelated with different activity patterns, reinforcing the princples of pattern language on the link between form and function (Alexander et al., 1977; Jacobs, 1961; Crooks et al., 2015). The clustering parameters employed capture distinct urban rhythms, though some granular rhythms may be lost. For example, while the selected parameters effectively identified activities in the Northern Quarter and Deansgate for eating out, smaller or transient activities may not have been fully captured. Exploring and experimenting with alternative parameters could help uncover rhythms at different spatial and temporal resolutions.

The rhythms analysed in this study align with observed urban activities during the Covid-19 pandemic, revealing significant disruptions and adaptations in Manchester City Centre. Social activities, such as "Bar/Pub Visits", "Cafes", and "Music and Nightlife", experienced sharp declines during lockdowns due to restrictions on gatherings and the closure of non-essential venues. Recovery for these activities was slow, reflecting the prolonged impact of social distancing measures and public caution in resuming pre-pandemic behaviours. Conversely, "Outdoor Leisure", recovered more quickly during easing phases, indicating a preference for safer, open-air environments. Digital engagement through "Social Media Activity" and "Mass Media" remained stable or increased, showcasing the growing importance of digital interactions as mobility was restricted. Other activities, including "Work-Related Activities" and "Shopping", demonstrated adaptability, with stability or incremental recovery during reopening phases, reflecting the adaptive dynamics of urban life under Covid-19 restrictions. Such insights could help inform new adaptive policies for urban planning (Rauws & De Roo, 2016).

Linking urban form to spatiotemporal activity patterns provided an alternative approach to understanding cities as CAS (Holland, 1992; Portugali, 2006). Through this lens, urban activities were seen as emergent behaviours shaped by bottom-up interactions, with and within urban form, influenced by top-down lockdown measures. This methodological flexibility could be extended to explore hierarchical and relational aspects of urban systems, such as nested patterns by incorporating network analysis and relational data approaches (Burgos-Thorsen & Munk, 2023; Ricci et al., 2018). Further advancement of this methodology to integrate urban form metrics to better capture urban typologies and characteristics would enhance the framework, enabling it to address broader and more complex questions related to interconnected and nested patterns within cities for design (Alexander et al., 1977).

5.3. Broader framework applicability and future methodological considerations

The tool and GUI interface demonstrated serves as a platform for data experimentation and interrogation. This aligns with research that calls for experimental methodologies to uncover alternative understandings of cities (Madsen, 2023). The GUI's ability to visualise spatiotemporal patterns provides researchers and planners with new analytical capabilities in capturing dynamics through complex spatiotemporal patterns, enabling the exploration of urban rhythms and their relationship with different urban form typologies.

The clustering parameters used in this study were tailored to capture daily rhythms and short-term spatiotemporal dynamics. To expand the framework, alternative configurations could explore longer temporal scales, such as seasonal variations or extended activity trends. Similarly, finer temporal parameters could reveal within-day rhythms, offering insights into transient activities. Integrating a time-geography framework (Hägerstrand, 1970) could further enhance the analysis by elucidating the flow and movement of activities across space and time. Such methodological developments would enable the framework to adapt to diverse research applications.

Data and API access limitations presents a growing challenge for researchers (Bruns, 2019). The shift towards paid APIs and restricted access to social media data necessitates innovative approaches, such as integrating data from multiple platforms or leveraging archival datasets. These methods could mitigate data constraints while expanding the scope for research. The methodology developed here also highlights the importance of methodological decisions, such as excluding bots in this paper, to ensure alignment with the study's focus. While bots can provide insights into information flows (Marres, 2015), their inclusion in understanding urban dynamics would skew analyses focused on complex social interactions. Future work could also explore topic weighting approaches.

Future research should extend the framework presented to incorporate stakeholder engagement, aligning with participatory approaches (Venturini et al., 2019) and participatory data design (Jensen et al., 2021). Involving urban planners, community groups, and policymakers in the design and analysis process could enhance the framework's practical relevance, and foster co-produced knowledge. This collaborative approach bridges the gap between computational possibilities and local insights, enabling alternative inclusive and impactful possibilities.

6. Conclusion

This paper introduced a novel framework for using LBSM, specifically geolocated tweets, to analyse urban activities as complex spatiotemporal patterns. By reviewing different LBSM data sources and their research use, the paper presented the framework and tool developed to uncover spatiotemporal dynamics and explore their relationship with urban form.

As demonstrated, the novel framework provided new analytical outcomes as patterns capturing rhythms, counts, distribution, sizes, and locations of urban activities. The methodology developed was tested in Manchester City Centre as an exploratory test case. When examined in the context of top-down and bottom-up dynamics, the spatiotemporal pattern outcomes offered new insights, such as how urban characteristics support different activities across different parts of Manchester City Centre. The results were presented in the context of Covid-19 lockdown measures as a significant top-down disruptor, and revealed bottom-up adaptation patterns over time, enabling the study of cities as CAS. The results include the identification of emergent collective behaviours, with urban activity patterns examined through temporal fluctuations, spatial distributions, and contextual influences. For example, fluctuations in activities such as "Music and Nightlife" and "Cafes and Desserts" highlighted the sensitivity of city dynamics to changes in policy and social behaviour. By demonstrating pattern results, the proposed framework advances our ability to study cities as hybrid urban-digital spaces.

Future research could extend this methodology for comparative analysis across different cities to identify unique patterns and different dynamics. Incorporating alternative LBSM data sources, such as geotagged photos, POI data, or demographic data could enhance relational understandings of activity patterns and interplay with urban form. The advancement of this framework could inform future research, but also be used in the context of urban planning, design and policy by providing actionable insights for adaptive and resilient cities.

CRediT authorship contribution statement

Mahmud Tantoush: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Ulysses Sengupta:** Writing – review & editing, Supervision, Resources. **Liangxiu Han:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank the leadership at Manchester Metropolitan University and the Manchester School of Architecture for supporting us at ARO\[CPU]Lab for their support. We are also grateful to the anonymous reviewers and the editor for their constructive feedback and valuable suggestions, which have greatly contributed to enhancing the quality of this paper

Data availability

The authors do not have permission to share data.

References

Alexander, C. (1979). *The timeless way of building* (Vol. 1). Oxford university press.
Alexander, C., Ishikawa, S., Silverstein, M., Jacobson, M., Fiksdahl-King, I., & Angel, S. (1977). *A pattern language: Towns, buildings, construction (not complete, from page 100 with notes)*. Oxford University Press.
Ash, J., Kitchin, R., & Leszczynski, A. (2018). Digital turn, digital geographies? *Progress in Human Geography, 42*(1), 25–43. https://doi.org/10.1177/0309132516664800
Ballatore, A., Rodgers, S., McLoughlin, L., & Moore, S. (2024). Facebook city: Place-named groups as urban communication infrastructure in Greater London. *Environment and Planning B: Urban Analytics and City Science, 0*(0), 1–19. https://doi.org/10.1177/23998083231224136
Batty, M. (2013). Big data, smart cities and city planning. *Dialogues in Human Geography, 3*(3), 274–279. https://doi.org/10.1177/2043820613513390
Batty, M. (2016). Big data and the city. *Built Environment, 42*(3), 321–337. https://doi.org/10.2148/benv.42.3.321
Birant, D., & Kut, A. (2007). ST-DBSCAN: An algorithm for clustering spatial-temporal data. *Data and Knowledge Engineering, 60*(1), 208–221. https://doi.org/10.1016/j.datak.2006.01.013
Bisanzio, D., Kraemer, M. U. G., Bogoch, I. I., Brewer, T., Brownstein, J. S., & Reithinger, R. (2020). Use of twitter social media activity as a proxy for human mobility to predict the spatiotemporal spread of COVID-19 at global scale. *Geospatial Health, 15*(1), 19–24. https://doi.org/10.4081/gh.2020.882
Blei, D., Jordan, M., & Ng, A. Y. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research, 3*, 993–1022. https://doi.org/10.1162/jmlr.2003.3.4-5.993
Boy, J. D., & Uitermark, J. (2016). How to study the city on instagram. *PLoS One, 11*(6), 1–16. https://doi.org/10.1371/journal.pone.0158161
Bruns, A. (2019). After the 'APIcalypse': Social media platforms and their fight against critical scholarly research. *Information Communication and Society, 22*(11), 1544–1566. https://doi.org/10.1080/1369118X.2019.1637447
Burgos-Thorsen, S., & Munk, A. K. (2023). Opening alternative data imaginaries in urban studies: Unfolding COVID place attachments through Instagram photos and computational visual methods. *Cities, 141*, Article 104470. https://doi.org/10.1016/j.cities.2023.104470
Chang, J., Boyd-Graber, J., Gerrish, S., Wang, C., & Blei, D. M. (2009). Reading tea leaves: How humans interpret topic models. In *Advances in neural information processing systems 22 - proceedings of the 2009 conference* (pp. 288–296).
Chauhan, U., & Shah, A. (2022). Topic modeling using latent Dirichlet allocation: A survey. *ACM Computing Surveys, 54*(7). https://doi.org/10.1145/3462478
Cheng, J., Han, L., & Jin, C. (2016). Big data for urban studies: Opportunities and challenges: A comparative perspective. In *2016 Intl IEEE conferences on Ubiquitous Intelligence & Computing, advanced and trusted computing, scalable computing and communications, cloud and big data computing, internet of people, and smart world congress* (pp. 1229–1234). https://doi.org/10.1109/UIC-ATC-ScalCom-CBDCom-IoP-SmartWorld.2016.0189
Cheng, T., & Wicks, T. (2014). Event detection using twitter: A spatio-temporal approach. *PLoS One, 9*(6), 1–10. https://doi.org/10.1371/journal.pone.0097807
Crooks, A. T., Croitoru, A., Jenkins, A., Mahabir, R., Agouris, P., & Stefanidis, A. (2016). User-generated big data and urban morphology. *Built Environment, 42*(3), 396–414. https://doi.org/10.2148/benv.42.3.396
Crooks, A. T., Croitoru, A., Stefanidis, A., & Radzikowski, J. (2013). #earthquake: Twitter as a distributed sensor system. *Transactions in GIS, 17*(1), 124–147. https://doi.org/10.1111/j.1467-9671.2012.01359.x
Crooks, A. T., Pfoser, D., Jenkins, A., Croitoru, A., Stefanidis, A., Smith, D., … Lamprianidis, G. (2015). Crowdsourcing urban form and function. *International Journal of Geographical Information Science, 29*(5), 720–741. https://doi.org/10.1080/13658816.2014.977905
De Roo, G., & Yamu, A. (2017). New ways of conditioning space and place in the dynamic and transformative environments: Liaising between the worlds of matter and the virtual. In C. Yamu, A. Poplin, O. Devisch, & G. De Roo (Eds.), *The virtual and the real in planning and Urban Design: Perspectives, practices and applications* (pp. 11–39). Routledge.
de Souza e Silva, A. (2006). From cyber to hybrid: Mobile technologies as interfaces of hybrid spaces. *Space and Culture, 9*(3), 261–278. https://doi.org/10.1177/1206331206289022
de Souza e Silva, A. (2009). Hybrid reality and location-based gaming: Redefining mobility and game spaces in urban environments. *Simulation and Gaming, 40*(3), 404–424. https://doi.org/10.1177/1046878108314643
de Souza e Silva, A., & Sheller, M. (Eds.). (2014). *Mobility and locative media: Mobile communication in hybrid spaces*. Routledge. https://doi.org/10.5617/jmi.v2i2.2378.
Ester, M., Kriegel, H., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the second international conference on knowledge discovery and data mining (KDD-96)* (pp. 226–231). AAAI Press.
Frias-Martinez, V., Soto, V., Hohwald, H., & Frias-Martinez, E. (2012). Characterizing urban landscapes using geolocated tweets. In *2012 ASE/IEEE international conference on privacy, security, risk and trust and 2012 ASE/IEEE international conference on social computing, SocialCom/PASSAT 2012* (pp. 239–248). https://doi.org/10.1109/SocialCom-PASSAT.2012.19
Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management, 35*(2), 137–144. https://doi.org/10.1016/j.ijinfomgt.2014.10.007
Gao, S., Janowicz, K., Montello, D. R., Hu, Y., Yang, J., Mckenzie, G., … Yang, J. (2017). A data-synthesis-driven method for detecting and extracting vague cognitive regions. *International Journal of Geographical Information Science, 31*(6), 1245–1271. https://doi.org/10.1080/13658816.2016.1273357
Goodchild, M. F. (2007). Citizens as sensors: The world of volunteered geography. *Geojournal, 69*(4), 211–221. https://doi.org/10.1007/s10708-007-9111-y
Hägerstrand, T. (1970). What about people in regional science? *Ninth European Congress of the Regional Science Association, 24*, 7–21. https://doi.org/10.1007/BF01954291
Haklay, M. (2013). Citizen science and volunteered geographic information: Overview and typology of participation. In D. Sui, S. Elwood, & M. Goodchild (Eds.), *Crowdsourcing geographic knowledge* (pp. 105–122). Springer.
Halegoua, G. (2020). The Digital City: Media and the social production of place. *New York University Press.*. https://doi.org/10.15826/csp.2020.4.2.100

Halegoua, G., & Polson, E. (2021). Exploring 'digital placemaking.'. *Convergence, 27*(3), 573–578. https://doi.org/10.1177/13548565211014828

Hochman, N., & Manovich, L. (2013). Zooming into an Instagram City: Reading through the local social media. *First Monday, 18*(7). https://doi.org/10.5210/fm.v19i7.4901

Holland, J. H. (1992). Complex adaptive systems. *Daedalus, 121*(1), 17–30.

Hu, Y., Gao, S., Janowicz, K., Yu, B., Li, W., & Prasad, S. (2015). Extracting and understanding urban areas of interest using geotagged photos. *Computers, Environment and Urban Systems, 54*, 240–254. https://doi.org/10.1016/j.compenvurbsys.2015.09.001

Hu, Y., & Wang, R. (2020). Understanding the removal of precise geotagging in tweets. *Nature Human Behaviour, 4*(12), 1219–1221. https://doi.org/10.1038/s41562-020-00949-x

Huang, H., Yao, X. A., Krisp, J. M., & Jiang, B. (2021). Analytics of location-based big data for smart cities: Opportunities, challenges, and future directions. *Computers, Environment and Urban Systems, 90*, Article 101712. https://doi.org/10.1016/j.compenvurbsys.2021.101712

Huang, X., Li, Z., Jiang, Y., Li, X., & Porter, D. (2020). Twitter reveals human mobility dynamics during the COVID-19 pandemic. *PLoS One, 15*(11), Article e0241957. https://doi.org/10.1371/journal.pone.0241957

Huang, Y., Li, Y., & Shan, J. (2018). Spatial-temporal event detection from geo-tagged tweets. *ISPRS International Journal of Geo-Information, 7*(4), 150. https://doi.org/10.3390/ijgi7040150

Institute for Government. (2022). *Timeline of UK government coronavirus lockdowns and measures, March 2020 to December 2021*. Retrieved from: Institute for Government Analysis https://www.instituteforgovernment.org.uk/data-visualisation/timeline-coronavirus-lockdowns.

Iranmanesh, A., & Alpar Atun, R. (2022). Reading the changing dynamic of urban social distances during the COVID-19 pandemic via Twitter. *European Societies, 23*(sup1), S872–S886. https://doi.org/10.1080/14616696.2020.1846066

Jacobs, J. (1961). *The death and life of great American cities.* Vintage Books.

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent Dirichlet allocation (LDA) and topic modeling: Models, applications, a survey. *Multimedia Tools and Applications, 78*(11), 15169–15211. https://doi.org/10.1007/s11042-018-6894-4

Jenkins, A., Croitoru, A., Crooks, A. T., & Stefanidis, A. (2016). Crowdsourcing a collective sense of place. *PLoS One, 11*(4), 1–20. https://doi.org/10.1371/journal.pone.0152932

Jensen, T. E., Birkbak, A., Madsen, A. K., & Munk, A. K. (2021). Participatory data design: Acting in a digital world. In G. Downey, & T. Zuiderent-Jerak (Eds.), *Making & doing: Activating STS through knowledge expression and travel* (pp. 117–136). MIT Press. https://doi.org/10.7551/mitpress/11310.001.0001.

Jiang, B., & Ren, Z. (2018). Geographic space as a living structure for predicting human activities using big data. *International Journal of Geographical Information Science, 33* (4), 1–16. https://doi.org/10.1016/B978-0-12-812970-8.00004-X

Kisilevich, S., Krstajic, M., Keim, D., Andrienko, N., & Andrienko, G. (2010). Event-based analysis of people's activities and behavior using Flickr and Panoramio geotagged photo collections. *Proceedings of the International Conference on Information Visualisation*, 289–296. https://doi.org/10.1109/IV.2010.94

Kitchin, R., & Dodge, M. (2011). *Code/space: Software and everyday life*. MIT Press. https://doi.org/10.7551/mitpress/9780262042482.001.0001

Kovacs-Gyori, A., Ristea, A., Havas, C., Resch, B., & Cabrera-Barona, P. (2018). #London2012: Towards citizen-contributed urban planning through sentiment analysis of twitter data. *Urban Planning, 3*(1), 75. https://doi.org/10.17645/up.v3i1.1287

Kuo, C.-L., Chan, T.-C., Fan, I.-C., & Zipf, A. (2018). Efficient method for POI/ROI discovery using Flickr geotagged photos. *ISPRS International Journal of Geo-Information, 7*(3), 121. https://doi.org/10.3390/ijgi7030121

Lansley, G., & Longley, P. A. (2016). The geography of twitter topics in London. *Computers, Environment and Urban Systems, 58*, 85–96. https://doi.org/10.1016/j.compenvurbsys.2016.04.002

Li, L., Goodchild, M. F., & Xu, B. (2013). Spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr. *Cartography and Geographic Information Science, 40* (2), 61–77. https://doi.org/10.1080/15230406.2013.777139

Liu, Y., Liu, X., Gao, S., Gong, L., Kang, C., Zhi, Y., & Chi, G. (2015). Social sensing: A new approach to understanding our socioeconomic environments. *Social Sensing: ANew Approach to Understanding Our Socioeconomic Environments, 105*(3), 512–530. https://doi.org/10.1080/00045608.2015.1018773

Longley, P. A., & Adnan, M. (2016). Geo-temporal Twitter demographics. *International Journal of Geographical Information Science, 30*(2), 369–389. https://doi.org/10.1080/13658816.2015.1089441

Longley, P. A., Adnan, M., & Lansley, G. (2015). The geotemporal demographics of twitter usage. *Environment and Planning A, 47*(2), 465–484. https://doi.org/10.1068/a130122p

Lukka, K. (2003). The constructive research approach. In L. Ojala, & O.-P. Hilmola (Eds.), *In case study research in logistics* (pp. 83–101). Turku School of Economics and Business Administration.

Madsen, A. K. (2023). Digital methods as ' experimental a priori ' – How to navigate vague empirical situations as an operationalist pragmatist. *Convergence, 39*(1), 1–22. https://doi.org/10.1177/13548565221144260

Madsen, A. K., Grundtvig, A., & Thorsen, S. (2022). Soft City sensing: A turn to computational humanities in data-driven urbanism. *Cities, 126*, Article 103671. https://doi.org/10.1016/j.cities.2022.103671

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. J., & McClosky, D. (2014). The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd annual meeting of the Association for Computational Linguistics: System demonstrations* (pp. 55–60).

Marres, N. (2015). Why map issues? On controversy analysis as a digital method. *Science Technology and Human Values, 40*(5), 655–686. https://doi.org/10.1177/0162243915574602

Martí, P., Serrano-Estrada, L., & Nolasco-Cirugeda, A. (2019). Social media data: Challenges, opportunities and limitations in urban studies. *Computers, Environment and Urban Systems, 74*, 161–174. https://doi.org/10.1016/j.compenvurbsys.2018.11.001

Martín, Y., Cutter, S. L., & Li, Z. (2020). Bridging twitter and survey data for evacuation assessment of Hurricane Matthew and Hurricane Irma. *Natural Hazards Review, 21* (2), 04020003. https://doi.org/10.1061/(asce)nh.1527-6996.0000354

Mckenzie, G., Janowicz, K., Gao, S., Yang, J., & Hu, Y. (2015). POI Pulse: A Multi-Granular , Semantic Signatures-Based Information Observatory for the Interactive Visualization of Big Geosocial Data. *Cartographica The International Journal for Geographic Information and Geovisualization, 50*(2), 71–85.

McLean, J. (2020). *Changing digital geographies: Technologies, environments and people*. Palgrave Macmillan. https://doi.org/10.1007/978-3-030-28307-0

Mislove, A., Lehmann, S., Ahn, Y.-Y., Onnela, J., & Rosenquist, J. N. (2011). Understanding the demographics of Twitter users. *Proceedings of the Fifth International AAI Conference on Weblogs and Social Media*, 554–557.

Newman, D., Lau, J. H., Grieser, K., & Baldwin, T. (2010). Automatic evaluation of topic coherence. In *Human language technologies: The 2010 annual conference of the North American chapter of the association for computational linguistic* (pp. 100–108).

Nguyen, Q. C., Li, D., Meng, H.-W., Kath, S., Nsoesie, E., Li, F., & Wen, M. (2016). Building a National Neighborhood Dataset from Geotagged Twitter Data for indicators of happiness, diet, and physical activity. *JMIR Public Health and Surveillance, 2*(2), Article e158. https://doi.org/10.2196/publichealth.5869

Niu, H., & Silva, E. A. (2020). Crowdsourced data Mining for Urban Activity: Review of data sources, applications, and methods. *Journal of Urban Planning and Development, 146*(2). https://doi.org/10.1061/(asce)up.1943-5444.0000566

Niu, H., & Silva, E. A. (2023). Understanding temporal and spatial patterns of urban activities across demographic groups through geotagged social media data. *Computers, Environment and Urban Systems, 100*, Article 101934. https://doi.org/10.1016/j.compenvurbsys.2022.101934

OpenStreetMap contributors. (2019). Planet dump. retrieved from https://planet.osm.org https://www.openstreetmap.org.

Osorio-Arjona, J., & García-Palomares, J. C. (2019). Social media and urban mobility: Using twitter to calculate home-work travel matrices. *Cities, 89*, 268–280. https://doi.org/10.1016/j.cities.2019.03.006

Portugali, J. (2006). Complexity theory as a link between space and place. *Environment and Planning A, 38*(4), 647–664. https://doi.org/10.1068/a37260

Quercia, D., & Saez, D. (2014). Mining urban deprivation from foursquare: Implicit crowdsourcing of city land use location. *IEEE Pervasive Computing, 13*(2), 30–36.

Rauws, W., & De Roo, G. (2016). Adaptive planning: Generating conditions for urban adaptability. Lessons from Dutch organic development strategies. *Environment and Planning B: Planning and Design, 43*(6), 1052–1074. https://doi.org/10.1177/0265813516658886

Redhead, N. (2012). Greater Manchester urban historic landscape characterisation project (HLC). *Archaeology Data Service [distributor]*. https://doi.org/10.5284/1017434

Resch, B., Summa, A., Zeile, P., & Strube, M. (2016). Citizen-centric urban planning through extracting emotion information from twitter in an interdisciplinary space-time-linguistics algorithm. *Urban Planning, 1*(2), 114. https://doi.org/10.17645/up.v1i2.617

Resch, B., Usländer, F., & Havas, C. (2018). Combining machine-learning topic models and spatiotemporal analysis of social media data for disaster footprint and damage assessment. *Cartography and Geographic Information Science, 45*(4), 362–376. https://doi.org/10.1080/15230406.2017.1356242

Ricci, D., Brilli, A., & Tassi, R. (2018). Repurposing digital methods for human-centered design. DRS 2018 - Catalyst 2018, Limerick, Ireland. *Design Research Society*. https://doi.org/10.21606/drs.2018.308

Rowe, F., Calafiore, A., Arribas-Bel, D., Samardzhiev, K., & Fleischmann, M. (2023). Urban exodus? Understanding human mobility in Britain during the COVID-19 pandemic using Facebook data. *Population, Space and Place, 29*(1), Article e2637. https://doi.org/10.1002/psp.2637

Sengupta, U., Rauws, W. S., & de Roo, G. (2016). Planning and complexity: Engaging with temporal dynamics, uncertainty and complex adaptive systems. *Environment and Planning B: Planning and Design, 43*(6), 970–974. https://doi.org/10.1177/0265813516675872

Sengupta, U., Tantoush, M., Bassanino, M., & Cheung, E. (2020). The hybrid space of collaborative location-based mobile games and the city: A case study of Ingress. *Urban Planning, 5*(4), 358–370. https://doi.org/10.17645/up.v5i4.3487

Shu, C. (2019). *Twitter will remove precise location tagging in tweets, citing lack of use*. Tech Crunch. https://techcrunch.com/2019/06/18/twitter-will-remove-location-tagging-in-tweets-citing-lack-of-use/#:~:text=In an announcement today from,their precise location in Tweets.

Sievert, C., & Shirley, K. (2015). LDAvis: A method for visualizing and interpreting topics. *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*, 63–70. https://doi.org/10.3115/v1/w14-3110

Singleton, A. D., & Spielman, S. E. (2014). The past, present, and future of geodemographic research in the United States and United Kingdom. *Professional Geographer, 66*(4), 558–567. https://doi.org/10.1080/00330124.2013.848764

Smith, D., & Crooks, A. T. (2010). *From buildings to cities: Techniques for multi-scale analysis of urban form and function (UCL working paper series)*. https://doi.org/10.1103/PhysRevE.78.016110

Stefanidis, A., Crooks, A. T., & Radzikowski, J. (2013). Harvesting ambient geospatial information from social media feeds. *GeoJournal, 78*(2), 319–338. https://doi.org/10.1007/s10708-011-9438-2

Stehle, S., & Kitchin, R. (2020). Real-time and archival data visualisation techniques in city dashboards. *International Journal of Geographical Information Science, 34*(2), 344–366. https://doi.org/10.1080/13658816.2019.1594823

Steiger, E., Resch, B., & Zipf, A. (2016). Exploration of spatiotemporal and semantic clusters of Twitter data using unsupervised neural networks. *International Journal of Geographical Information Science, 30*(9), 1694–1716. https://doi.org/10.1080/13658816.2015.1099658

Tenney, M., Hall, G. B., & Sieber, R. E. (2019). A crowd sensing system identifying geotopics and community interests from user-generated content. *International Journal of Geographical Information Science, 33*(8), 1497–1519. https://doi.org/10.1080/13658816.2019.1591413

Venturini, T., Munk, A., & Meunier, A. (2019). Data-sprinting: A public approach to digital research. In C. Lury, P. Clough, M. Michael, R. Fensham, S. Lammes, A. Last, & E. Uprichard (Eds.), *Routledge handbook of interdisciplinary research methods* (pp. 158–163). Routledge. https://doi.org/10.4324/9781315714523-24.

Xing, H., & Meng, Y. (2018). Integrating landscape metrics and socioeconomic features for urban functional region classification. *Computers, Environment and Urban Systems, 72*, 134–145. https://doi.org/10.1016/j.compenvurbsys.2018.06.005

Yan, Y., Schultz, M., & Zipf, A. (2019). An exploratory analysis of usability of Flickr tags for land use/land cover attribution. *Geo-spatial Information Science, 22*(1), 12–22. https://doi.org/10.1080/10095020.2018.1560044

Yao, F., & Wang, Y. (2020). Tracking urban geo-topics based on dynamic topic model. *Computers, Environment and Urban Systems, 79*, Article 101419. https://doi.org/10.1016/j.compenvurbsys.2019.101419

Ye, C., Zhang, F., Mu, L., Gao, Y., & Liu, Y. (2021). Urban function recognition by integrating social media and street-level imagery. *Environment and Planning B: Urban Analytics and City Science, 48*(6), 1430–1444. https://doi.org/10.1177/2399808320935467

Yin, R. K. (2014). *Case study research: Design and methods* (5th ed.). Sage.

Yuan, N. J., Zheng, Y., Xie, X., Wang, Y., Zheng, K., & Xiong, H. (2015). Discovering urban functional zones using latent activity trajectories. *IEEE Transactions on Knowledge and Data Engineering, 27*(3), 712–725. https://doi.org/10.1109/TKDE.2014.2345405

Zhan, X., Ukkusuri, S. V., & Zhu, F. (2014). Inferring urban land use using large-scale social media check-in data. *Networks and Spatial Economics, 14*, 647–667. https://doi.org/10.1007/s11067-014-9264-4

Zhang, Y., & Eick, C. F. (2019). Tracking events in twitter by combining an LDA-based approach and a density-contour clustering approach. *International Journal of Semantic Computing, 13*(1), 87–110. https://doi.org/10.1142/S1793351X19400051