

# IDENTIFYING ANCHOR POINTS IN CROWD-SOURCED SOCIAL DATA

**Nick Malleson** and Mark Birkin  
University of Leeds, UK  
n.malleson06@leeds.ac.uk

## ABSTRACT

### *Introduction*

Part of the appeal of agent-based modelling is the ability to directly model the behaviour and interactions of the individual units from which larger system patterns emerge. However, this advantage is tempered by the difficulty of obtaining data at appropriate geographical and temporal scales on which to validate the behaviour of simulated individuals. This paper will present preliminary research that attempts to identify areas of particular interest for users of social network services based on the geographical locations from which they post messages. Specifically, we employ a technique borrowed from physical geography to identify peaks in landscape data to automatically process a corpus of thousands of individual users. It is intended that the results of this approach can be used to improve the accuracy of an agent-based model of individual daily activity patterns.

### *Background*

Crowd-sourced data have the potential transform our approaches to behavioural validation in agent-based modelling. Social networking services such as Facebook and Twitter, as well as commercial data such as supermarket loyalty schemes or mobile telephone locations, contain a vast amount of information about individual human activity patterns. Among other problems, a challenge for agent-based modellers – or social scientists who are simply interested in learning more about daily activity patterns – is to extract intelligence from otherwise noisy and biased data.

Figure 1 illustrates part of the challenge. It presents the locations and density of a single user's messages posted on Twitter with the GPS location enabled. Clear areas of high density emerge which could relate to areas that the user commonly frequents (e.g. 'anchor points'). However, defining the size and shape of an anchor point automatically is non-trivial, particularly for social network data which vary so considerably from one user to another.

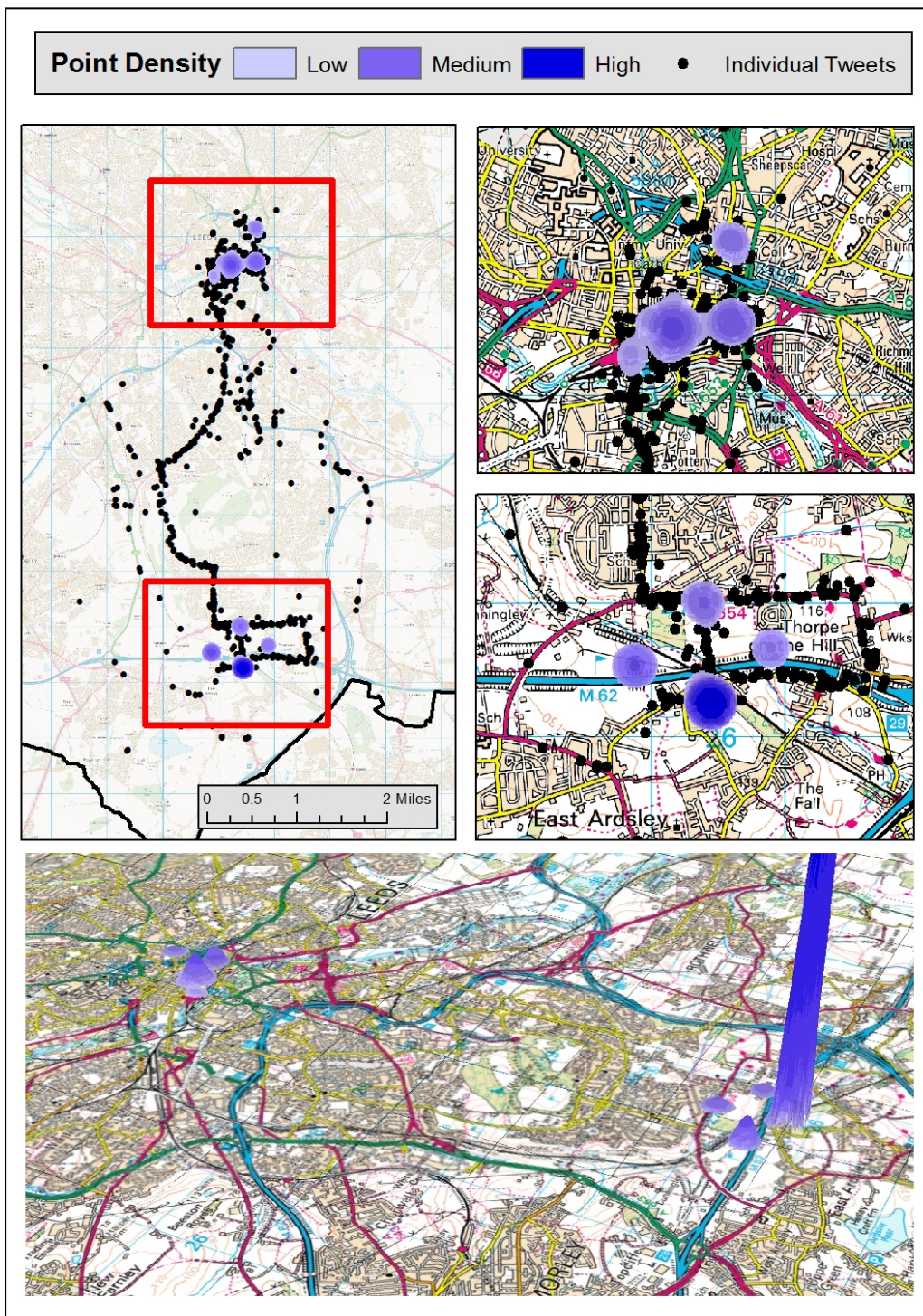
### *Method*

This paper will present the results of a preliminary approach to the identification of anchor points for thousands of distinct users who have posted messages on social media services. It uses the novel application of a peak-finding algorithm developed in physical geography to identify features in landscapes based on ground height and surrounding slopes. This method is applied to a message density landscape in order to identify places that have a higher spatial message density than their surroundings.

### *Results & Discussion*

Early results suggest that it is possible to identify the distinct areas that shape a person's activity space from the locations of messages posted to social media services. We present a range of indices that characterise the characteristics of different users' activity spaces.

Although in early stages, this work has clear implications for modelling behaviour in agent-based models. For example, these results could be used to inform the behaviour of individual people in a model of daily routine activities – either as a means of validating a model or to build new theories of behaviour on which the model will be based.



**Figure 1. The locations and density of a user's messages posted to Twitter.**

## KEYWORDS

Crowd-sourced data, ABM, Simulation