

@Twitter Tourism

Bartosz Hawelka¹, Peter Ranacher², Izabela Sitko³ and Thomas Lampoltshammer⁴

^{1,2,4} University of Salzburg, Doctoral College GIScience, Salzburg, Austria

³Jagiellonian University, Department of GIS, Cartography and Remote Sensing, Krakow, Poland

Today people produce and leave behind high numbers of digital fragments, which all together can be used to reason about collective behaviour. All these digital footprints come from a variety of different data probes. One very popular probe delivering data about society is the social networking service and microblog Twitter. The usefulness of Twitter as a data source for scientific research has already been proven (BOLLEN and MAO 2011, DODDS et al. 2011, GOLDBERGER and MACY 2011). However, only a few authors such as CHENG and SUI 2011 and NOULAS et al. 2011 try to explicitly take into consideration the spatial dimension of this data probe in their research. Our work contributes to this branch of research and analyses the spatial aspects of a society's behaviour – in our case tourism – with the help of Twitter data.

According to the statistics published by Twitter, the platform currently has more than 140 million active users who publish all together 340 million tweets per day. Even though geo-located tweets constitute less than 1% (with tendency to rise) of all the tweets, in terms of absolute numbers we consider them as substantial and representative to indicate spatial collective behaviour.

In order to analyse collective spatial behaviour, a very important first step usually is to gain insight into the distribution of the population in space, i.e. to find out those locations where the individual actors of the population have their centre of living. This means that also for the analysis of spatial behaviour of Twitter users it is an important prerequisite to find a user's location of most frequent occurrence. We refer to this location as a user's permanent location. Twitter offers its users the possibility to declare their permanent location. This declared location is, however, often far from a user's actual centre of living and therefore insufficient for geographic analysis. We argue that a Twitter user's permanent location is the one location where the user tweets most often and propose a methodology for finding it.

We introduce an adaptation of the well known DBSCAN clustering algorithm by Ester et al. 1996. This algorithm identifies natural clusters in a set of n-dimensional data by comparing the distances between each two points in a data set. As Twitter data are located in geographical space it is necessary to replace the – usually metric – distance function with a spherical one.

We apply our methodology to reason about the collective image of tourism in an urban environment. As a case study we choose the city of Salzburg as it is a substantial touristic hot spot, visited by thousands of tourists from all over the world. By having the knowledge about the permanent location for each Twitter user we are able to a) tell apart tweets from local residents and visitors, b) identify the visitors' locations of origin. The general aim of a study is to identify all the twitter user who visited Salzburg between January and April 2012, determine the country of residence and check them against the official statistic about tourism in the city of Salzburg. Thus, we assess the potential of the representativeness of Twitter data for analysing touristic phenomena.