

## SHARING INSIGHTS ON HATE CRIME: NEW METHODS AND FORMS OF DATA

### BACKGROUND

Understanding hate crime is a priority for police forces across England and Wales. Since the EU referendum in June 2016, there has been renewed emphasis on preventing hate crime and providing support for victims.

With this in mind, the *Sharing insights on hate crime: new methods and forms of data* project developed new relationships between academics at the Universities of Leeds and Liverpool, Lancashire Constabulary and the Safer Lancashire Partnership through which to drive improvements in policing and service provision for victims of religiously and racially motivated hate crime.

The research was based on the quantitative and spatial analysis of secondary data sources. The project exchanged learning between academic and policing partners in methods for examining Twitter data and routinely captured police data in novel ways, to develop a richer and more nuanced understanding of the dynamic and changing risks to communities.

### KEY FINDINGS

Key findings from the analysis of Twitter and routinely captured police data for hate crime in Lancashire include:

- It is possible to create an English language classifier to identify online hate speech on Twitter.
- 10% of the victims are university students and school pupils.
- Hate crime offenders and victims tend to live in separate neighbourhoods, with most of the hate incidents occurring in the town centre where the two populations meet.
- Hate crimes and incidents frequently occur at night (between 7pm and 6am); hate crime incidents frequently occur whilst victims are at work/in the workplace.

The research has provided Lancashire Constabulary with a clearer picture of the profile of victims and offenders of hate crime as well as the times and locations where the incidents happen. It has also emphasised the value of social media data when dealing with under-reported crimes.

Lancashire Constabulary has used this work to refocus a multi-agency response to tackling hate crime in Lancashire.

## THE STUDY

This project speaks directly to public and policing concerns regarding hate crime and community tensions in the wake of the EU referendum (June 2016) and managing the perceptions of and tensions between British and foreign nationals (from the EU and beyond) who are resident in the United Kingdom. It also links directly to the current work of Lancashire Constabulary aimed at understanding levels of hate crime.

The project was born out of co-investigators roles in the N8 Policing Research Partnership (N8PRP), and specifically the work of the Data Analytics activity strand thereof. The N8PRP facilitated relationships and dialogue between the co-investigators — Lightowlers and Malleson — and representatives from Lancashire Constabulary and the Safer Lancashire Partnership from which this project developed.

Securing an Impact Acceleration Account grant from the Leeds Social Sciences Institute (LSSI) as well as internship from the Leeds Institute of Data Analytics (LIDA), academics from the Schools of Law and Geography at the University of Leeds<sup>1</sup>, worked together with Lancashire Constabulary and Safer Lancashire to tackle these concerns by exploring new sources of data and methods for interrogating these. Impact Acceleration Account grants are specifically designed to support researchers in the social sciences to connect and engage with external organisations to maximise the impact and influence of their research on society and the economy.

*Sharing insights on hate crime: new methods and forms of data* shared learning about new forms of data and analytical techniques for gauging the ‘temperature’ of hate in local communities. This learning ensued based on the analysis of data on reported incidents of hate crime, area demographics and voting patterns (both in the EU referendum and the last general election in the United Kingdom) as well as geo-coded Twitter data.

The project’s primary outcome was to promote and increase the analytical capacity and skills of community safety and police analysts within partner organisations by enhancing their knowledge and skills in using new forms of data and methods of analysis when conducting analyses of hate crime. A secondary aim was to explore the value of Twitter data for identifying incidents of hate.

This knowledge exchange project has equipped community safety and police analysts with additional skills to illicit insights into the dynamic and changing risks to communities and profiles of hate crime perpetrators and victims. It did so by developing new relationships between academics at the University of Leeds, University of Liverpool, Lancashire Constabulary and the Safer Lancashire Partnership. As a consequence, it introduced policing analysts to a rich and wider range of complimentary data sources and methods for developing a more nuanced understanding of hate crime.

In turn, this enriched and informed Lancashire Constabulary’s strategic and partnership intelligence assessments, which are ultimately aimed at:

- promoting of the issue of hate crime as important amongst policing partners;
- supporting the targeting of finite resources to areas and populations of greatest risk;
- improving local services and safeguarding of vulnerable communities;
- encouraging the public to report incidents of hate crime and encourage victims to come forward; and
- assisting Her Majesty’s Inspectorate of Constabulary (HMIC) in identifying barriers to reporting and how police can better respond to incidents of hate crime.

Linked with the project aim of identifying new forms of data and analytical techniques for understanding hate crime, the project identified two key work packages to help further understand the profile of hate

incidents and crimes in Lancashire. In the first, Chenevoy (LIDA intern) explored the potential of making use of Twitter data with which to identify incidents on hate online. In the second, Chenevoy examined the profile of recorded hate incidents and the characteristics of victims and offenders as well as the locations of the incidents.

### **Application of Natural Language Processing to identify online hate on Twitter**

A core objective of this strand of the research was to investigate whether online hate on Twitter could be used as a proxy for 'real life' hate happening in Lancashire. The ambition of this part of the project was to enable Lancashire Constabulary to harness new forms of social media data (Twitter) for their own analysis of hate crime in the area by applying machine learning.

The research was based on the spatial analysis of tweets sent by people in Lancashire during the study period (December 2015 to February 2017). These tweets were identified based on both the home town displayed on the profile of Twitter users and the precise geo-tags of where the tweets were sent from. In total, 1,246,918 tweets with a home town location and 389,410 tweets with a geo-tag were collected within the boundaries of Lancashire.

Chenevoy used the open source software Python to develop a machine learning algorithm that was able to capture hate speech from Twitter data. This first involved training the machine learning algorithm with tweets that had been manually classified. The accuracy of the trained algorithm was then tested on new tweets for which the classification outcome was already known, using a 10-fold cross-validation method. Spambots (autonomous program designed to publish tweets automatically) were also identified and tweets sent by these removed before analysis.

Once the classifier had shown a reliable level of accuracy, it was used to identify those Lancashire tweets containing hateful speech.

- A total of 98 geo-tagged tweets were found to contain hateful speech amongst the 194,574 geotagged tweets in Lancashire.
- A total of 3,903 tweets with home town location were found to contain hateful speech amongst the 1,189,066 tweets with home town location in Lancashire.

The identified hateful tweets were displayed on density maps at county and street level.

Whilst, the limited number of tweets with geo-tags identified as hateful by the classifier does not allow for as meaningful a geographical interpretation when considered at individual town level, this work revealed that it is possible to create an English language classifier which accurately identifies online hate speech on Twitter.

As such, the learning from this project has emphasised the valuable information provided by social media data when dealing with under-reported crimes. Twitter produces real time data which can be helpful in generating a spatial and temporal 'temperature check' of different localities. The algorithm developed in this project offers the potential to be used by Lancashire Constabulary to monitor levels of hate, thus ensuring resources can be allocated effectively to respond to emerging community tensions.

### **Analysis of police-recorded hate crime in Lancashire**

The objective of this second strand of the research was to describe the nature as well as spatial density of reported hate incidents and crimes in Lancashire by making sense of the police-recorded data. This was based on the quantitative and spatial analysis of administrative data provided by Lancashire Constabulary.

The data comprised the 6,485 police-recorded hate crimes and incidents which took place in Lancashire during the study period (December 2015 to February 2017), which were shared with academic partners subject to secure data transfer and storage arrangements and appropriate data sharing agreements.

In addition to performing basic descriptive statistics to summarise the data, density maps were produced for various towns of Lancashire. These maps highlight hotspots of hate incidents, as well as areas populated by a high proportion of victims or offenders.

Some of the key findings from the analysis of routinely-captured police data in Lancashire include:

- The majority of the employed victims work in public facing roles with most of these being taxi drivers (10%) and shop employees (9.5%). This suggests that hate incidents can frequently occur in the workplace/whilst on duty at work.
- 10% of the victims are university students and school pupils.
- 85% of hate crimes and incidents occur at night (between 7pm and 6am).
- Hate crime offenders and victims tend to live in separate neighbourhoods, with most of the hate incidents occurring in the town centre where the two populations meet.

Despite efforts deployed by the police to encourage the reporting of hate crime, under-reporting remains a significant problem. As such, it is crucial to make sense of the crimes that *are* indeed reported to the police. This research provided Lancashire Constabulary with a clearer picture of the profile of victims and offenders of hate crime as well as where and when the incidents happen (schools and town centre), thus ensuring resources can be allocated effectively.

The learning from this project has been used to re-focus a multi-agency response in tackling hate crime in Lancashire. It has also emphasised the importance of accurate crime recording and application of appropriate analytical techniques.

### **IMPLICATIONS FOR FURTHER RESEARCH**

The project's key output was the development of an algorithm for identifying hateful incidents in textual social media data. This allows the Constabulary to monitor levels of hate in the locality, thus ensuring a response can be deployed effectively before there is an escalation of harm. The algorithm and its source code (<https://github.com/mednche/Hate-Crime-Project>) was accompanied by an internal methodological report for Lancashire Constabulary's own ongoing use and further development.

The project developed the project team's skills in using new data sets and methods which can later be transferred to other areas of strategic priority or other crime 'problems' both in Lancashire and beyond, as similar learning can be adopted and applied in other police services or potentially be scaled up and applied nationally. The algorithm can also be adapted and modified for use on other sources of textual and social media data.

**Authors:** Carly Lightowlers,<sup>1</sup> Natacha Chenevoy,<sup>2</sup> Nicholas Malleson,<sup>2</sup> S. Beeley,<sup>2</sup> F. Blair,<sup>2</sup> Scott Keay,<sup>3</sup> R. Bretherton,<sup>3</sup> K. Stone,<sup>3</sup> R. Eckersley,<sup>3</sup> D. Chapman<sup>3</sup> and F. Pascale<sup>3</sup>

1. University of Liverpool;      2. University of Liverpool;      3. Lancashire Constabulary

**Further Information:** A full project report can be found [here](#).