

Using Geotagged Twitter Data to Uncover Hidden Church Populations

Anthony-Paul Cooper

Introduction

Alongside the narrative of church decline, as set-out by proponents of the secularisation thesis¹, recent research has presented compelling evidence of the existence of pockets of church growth in the UK.² While such evidence of church growth is not necessarily mutually exclusive with the secularisation thesis – which does not require church decline to be universal, even, or rapid^{3 4} - it has driven an increased research focus on church growth and the secularisation thesis more generally. In responding to this increased focus on church growth research, Bruce⁴ conveyed the importance of researchers working towards understanding whether or not the church is experiencing net growth or net decline as well as how such net growth or net decline compares to wider population changes in the location being studied. A number of challenges mean that it is currently not possible to robustly identify the net status of church attendance. Such challenges include, for example, conflicting measures of church attendance⁵, the comparability of data captured from different studies and locations and the difficulty associated with researchers capturing data from congregations which regularly form, merge, move, split and close.

The main aim of this chapter is to investigate whether or not Twitter data – known as tweets⁶ – can be used to identify churches which were previously not identified for inclusion in quantitative research. Should such data allow the identification of 'hidden churches', this would enable the research method to be re-used in future studies in order to reduce under counting of church attendance and ensure that such quantitative research had as broad a coverage as possible. Such improvements in church growth data would help the research community to move closer towards understanding whether or not the church is experiencing net growth or net decline. This research builds on the findings of previous studies which have investigated church-related tweets by further exploring the possible usefulness of Twitter data to sociologists of religion. The use of such data presents a unique opportunity to researchers to combine freely and quickly available

¹ See, for example: C. Brown, *The Death of Christian Britain*, (London: Routledge, 2001); S. Bruce, *God is Dead: Secularization in the West*, (Oxford: Wiley-Blackwell, 2002).

² See, for example: P. Brierley, *Capital Growth*, (Tonbridge: ADBC Publishers, 2013); D. Goodhew, 'Church growth in Britain, 1980 to the present day' in D. Goodhew (ed.), *Church Growth in Britain 1980 to the Present*, (Farnham: Ashgate, 2012): 3-20.

³ S. Bruce, 'Secularization and church growth in the United Kingdom', *Journal of Religion in Europe*, 6, (2013): 273-296.

⁴ S. Bruce, 'Further thoughts on church growth and secularization', *Journal of Religion in Europe*, 6, (2013): 316-320.

⁵ D. Goodhew, 'Church growth in Britain: A response to Steve Bruce', *Journal of Religion in Europe*, 6, (2013): 297-315.

⁶ At the time of the study, 140 character entries posted by users of the social media platform Twitter. Since the study, Tweets have been extended and can now contain up to 280 characters.

online data with more traditional research methods and data sources in order to enhance their understanding of the church and of church attendees.

Church growth in London

The 2011 census⁷ revealed that the usually resident population of the UK's capital on census day (27th March 2011) was 8.17 million – a 14 percent increase compared to the previous census 10 years earlier, returning the population of Greater London to a level similar to that recorded in the 1951 census – 8.16 million.

The 2011 census⁸ also revealed that there were 3.96 million London residents who described themselves as 'Christian' – a decrease, as was the case in all regions in England and Wales, between the 2001 and 2011 census.

The London Church Census⁹, conducted in 2012 by Peter Brierley, provides further context and detail around church activity in London, and probably represents the most comprehensive source of data on the church in London. The Church Census took place on 14th October 2012 and found that an estimated 720,000 people (approximately 9% of the population) attended around 4,800 churches on a Sunday. The study reported that the number of churches had increased from 4,100 (17% growth) between 2005 and 2012, while the number of church attenders had increased from 620,000 (16% growth) between 2005 and 2012. In both cases, the rate of growth observed took place at a faster rate than elsewhere in England.

The study noted that around 27% of London's churches or congregations were black majority churches or immigrant churches. The Church Census also noted that while approximately 9% of London's population attended church on a Sunday, this ranged from around 8% of London's white population to around 19% of the black population. These findings were consistent with a number of prior studies which had reported a relationship between church growth and immigration in the UK¹⁰.

It is clear that the picture of Christianity in London is complex, as is the case elsewhere. While the number of individuals who describe themselves as 'Christian' has been seen to decrease, the number of churches and of church attenders has been seen to increase. Furthermore, the number of church attenders in London has been observed to increase faster than the wider population of London. From the evidence available however, it is not possible to identify whether the same is true of the migrant population attending church in London. Furthermore, due to the fact that churches regularly form, merge, move, split and close, it is difficult to establish the extent to which church statistics related to London might under count the number of churches or church attenders, through missing information regarding churches which could be included in quantitative research studies. The London Church Census utilised a snowballing method where participating churches were encouraged to respond to the researcher with details of other churches which they are aware had come into being. While this represents probably the most effective method of identifying churches using traditional research methods, some churches will

⁷ Census Information Scheme GLA Intelligence, 2011 Census first results: London boroughs' population by age by sex (<https://files.datapress.com/london/dataset/2011-census-demography/2011-census-first-results.pdf> last accessed 30 August 2016, 2012).

⁸ Office for National Statistics, Census gives insights into characteristics of London's population (http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/dcp29904_291554.pdf last accessed 30 August 2016, 2012).

⁹ Brierley, *Capital Growth*.

¹⁰ See, for example: C. Marsh, 'The diversification of English Christianity: The example of Birmingham', in Goodhew (ed.), *Church growth in Britain*: 193-206; H. Osgood, 'The rise of black churches', in Goodhew (ed.), *Church growth in Britain*: 107-126.

inevitably slip through the net, leading to some degree of under counting. This does not undermine the findings of the London Church Census – which necessarily represent a conservative estimate of the church in London – but does highlight the role which technology and in particular, innovative online research methods, might play in complementing traditional offline research methods – in this case by highlighting the locations of churches which might otherwise be excluded from survey research.

Twitter data

With 313 million monthly active users¹¹ posting hundreds of millions of tweets each day, Twitter is a copious and valuable source of social data. The value of Twitter data is enhanced by the ease with which it can be accessed – by purchasing complete data sets from third party re-sellers¹² or by generating samples of freely available data using Twitter's own Application Programming Interface (API)¹³. As such Twitter offers the research community the opportunity to understand the behaviours and activities of a diverse community of individuals using a range of qualitative and quantitative research methods. When obtaining Twitter data, as well as receiving the tweet content, researchers are additionally presented with metadata which can be included in analyses, and are often as valuable, if not more valuable, than the actual tweet content. These metadata include, for example, the username of the individual who posted the tweet, the date and time when the tweet was posted and, in some cases, geodata showing the location from which the tweet was posted – expressed as a latitude and longitude value. The research presented in this chapter seeks to investigate whether or not Twitter metadata, in this case geodata, can be used to identify the locations of churches within London.

Previous research studies have explored the use of use of Twitter geodata to answer a range of research questions. Cheng, Caverlee and Lee¹⁴, for example, proposed and evaluated a probabilistic approach to identifying the city-level location of a Twitter user, based only on the content of the user's tweets. On average, the research found that the probabilistic approach was able to place “51% of Twitter users within 100 miles of their actual location” (p759). The findings of this research provide encouragement to researchers that it might be possible, in future, to attribute locations to tweets which do not have associated geodata with some degree of accuracy. Furthermore, the findings of this paper contribute to social media research theory more generally, by demonstrating a relationship between the content of Twitter data and the locations from which tweets are posted.

Similarly, research by the GIScience Research Group in 2015¹⁵ analysed semantic and spatiotemporal clustered tweets, finding that they correlated strongly when compared to workplace census data, “being a good indicator and representative proxy for analysing workplace-based activity” (p255). This research provides further confidence in the geodata provided with tweets, and the usefulness of such data in a research setting.

¹¹ See: <https://about.twitter.com/company/> (last accessed 30 August 2016).

¹² Such as Gnip: <https://gnip.com/> (last accessed 29 August 2016).

¹³ See: <https://dev.twitter.com> (last accessed 20 August 2016).

¹⁴ Z. Cheng, J. Caverlee & K. Lee, ‘You are where you tweet: a content-based approach to geo-locating twitter users’, in proceedings of the 19th ACM International Conference on Information and Knowledge Management, Toronto, Canada, October, 2010: 759-768.

¹⁵ E. Steiger, R. Westerholt, B. Resch & A. Zipf, ‘Twitter as an indicator for whereabouts of people? Correlating Twitter with UK census data’, *Computers, Environment and Urban Systems*, 54, (2015): 255-265.

In 2011¹⁶, researchers presented an approach to utilising Twitter data in support of crisis management. The proposed approach made use of a map-based interactive web application, which made sense of tweets based on place, time and concept characteristics. The implications of this research were clear – that publicly available data could be harnessed and used in support of offline activity.

More recently, in their 2013 paper¹⁷ Walther and Kaisser demonstrate that it is possible to monitor all tweets from a particular geographic region, identify locations showing a high amount of activity and analyse these clusters of posts using machine learning in order to identify whether or not these clusters of high activity represent real-world events. The authors found that their method was able to do so with high precision and recall. This important research finding, that event-detection is possible using geotagged Twitter data, provides a theoretical basis on which to assume that it might also be possible to identify community and place locations using geotagged Twitter data – an assumption which the research presented in this chapter seeks to robustly address.

Frias-Martinez and Frias-Martinez¹⁸ provide further illustration of the potential benefits of geolocated Twitter data through their case studies of data posted from Manhattan (USA), London (UK) and Madrid (Spain), in which they demonstrated that unsupervised learning could be used to determine urban land use by clustering geographical locations with similar tweeting activity. The authors posited that geolocated tweets could therefore be used as a valuable source of data to assist urban planning applications. The research presented in this chapter similarly seeks to further this understanding of the potential usefulness of social media data, by identify possible uses which could be made of geolocated Twitter data by sociologists of religion.

Twitter and the church

While a number of studies, such as those mentioned previously in this chapter, have considered the use of geolocated Twitter data to answer a broad range of research questions, relatively few studies to date have focused on the use of social media data in research projects investigating the church. This chapter builds on research previously conducted by the author using church-related tweets. In a small-scale thematic analysis¹⁹ of 108 tweets posted from within the London Borough of Camden over a 10 week period, research identified the topics which Twitter users tended to discuss in church-related tweets, as shown in table 6.1.

¹⁶ A. M. MacEachren, A. C. Robinson, A. Jaiswal, S. Pezanowski, A. Savalyev, J. Blanford P. & Mitra, 'Geo-Twitter analytics: Applications in crisis management', in proceedings of the 25th International Cartographic Conference, Paris, France, 2011.

¹⁷ M. Walther & M. Kaisser, 'Geo-spatial event detection in the Twitter stream', in Proceedings of the 34th European Conference on Information Retrieval, ECIR, 2013: 356-367.

¹⁸ V. Frias-Martinez & E. Frias-Martinez, 'Spectral clustering for sensing urban land use using Twitter activity', *Engineering Applications of Artificial Intelligence*, 35, (2014): 237-245.

¹⁹ A. P. Cooper, 'Unwrapping Camden's church tweeters: A small-scale thematic study of Twitter data' in K. Woodfield (ed.), *Social media in social research: Blogs on blurring the boundaries*, (London: NatCen Social Research, eBook, 2014).

Table 1: Thematic coding labels used in a previous study to categorise church-related tweets from the London Borough of Camden, with the volume of tweets captured by each label. *Source*: Cooper, 2014²⁰.

Qualitative coding label	No. of tweets captured by code
Discussion about church attendance	59
Discussion about church service content	5
Discussion about theology, the Bible or belief	15
General church discussion	40
Miscellaneous	5
Non-church discussion	4
Unknown theme	10

As can be seen from table 6.1, the most frequent topic of conversation within the sample of tweets analysed (approximately 55%) was 'discussion about church attendance'. The study noted that in some cases tweets about church attendance even captured specific mention of the church from which the tweet was posted. Though this small-scale study could have benefited from a greater sample size, this particular finding was useful as it indicated that future studies, such as the research work presented in this chapter, may be able to use tweets posted from church locations to increase understanding and build a picture of church locations.

In a separate follow-up study²¹, the author analysed a sample of 1,004 English language tweets which contained the word *church* and geolocation data necessary to attribute them to a London borough. The analysis found that there was a statistically significant relationship between the net sentiment score²² of church-related tweets and the presence of church growth in the London borough from which the tweets were posted. The identification of this relationship confirmed that there exists a relationship between online discussion around the church and offline engagement with the church.

The findings of both of these prior studies point towards the rich insights which can be gained through the analysis of Twitter data by researchers interested in better understanding the church. When combined with more traditional research methods, analysis of this freely available source of data can enable sociologists of religion to better understand how people discuss the church online, how church attendees interact with each other, and the church, online and how changes in online discussion of the church may be related to changes in offline engagement with the church. One of the key advantages of Twitter data, and of social media data more generally, is the ease and speed with which it can be accessed and the rich metadata which is provided along with the post content. The research presented in this chapter seeks to advance previous research in this field by investigating whether or not analysis of church-related tweets containing geodata can be used to identify churches which were not included in previous quantitative studies of church attendance.

²⁰ Cooper, 'Unwrapping'.

²¹ A. P. Cooper, 'Assessing the possible relationship between the sentiment of church-related tweets and church growth', *Studies in Religion/Sciences Religieuses*, 46, (2017): 37-49.

²² A measure of whether a tweet contains mostly positive, neutral or negative content.

The research questions

The research presented in this chapter will consider whether or not previously hidden churches can be identified using the geodata contained in church-related tweets. As such the research will begin by investigating whether or not geolocated Twitter data can be used to identify church locations. Assuming this question can be answered affirmatively, any identified churches will be compared to the list of churches included in the 2012 London church census, in order to identify whether or not any identified churches were previously not included in that major research study. These research aims can be clearly defined as two research questions:

Research question 1: Can geolocated Twitter data be used to identify church locations?

Research question 2: Can geolocated Twitter data be used to uncover previously hidden churches?

In order to answer these questions, this study will seek to gather a collection of church-related tweets, posted from within London over a number of months. The findings of a previous study described earlier in this chapter, which investigated the content of church-related tweets posted within the London Borough of Camden²³, indicate that some of these tweets will have been posted from within church locations, while others will have been posted from a plethora of other locations. In order to identify possible church locations from the collection of tweets, the analysis will seek to group together clusters of tweets which appear to have been posted from the same location. These tweets can then be manually inspected and qualitatively coded based on whether or not they are likely to have been posted from a church location. Where clusters of tweets are likely to have been posted from a church location, further research can be conducted in order to verify the presence of a church in that location and to obtain a name and postcode for any identified churches. The names and postcodes of churches identified using Twitter data in this study can then be compared to the list of churches included in the 2012 London Church Census²⁴ in order to confirm whether or not any previously hidden churches have been identified. The method used to conduct this analysis is presented in more detail in the following section of this chapter.

Finding 'new' churches using social media

The Twitter API, was used to collect a sample of tweets posted in the English language each Sunday over a 23 week period, commencing 20th April 2014. Each tweet in the sample contained the word *church* and was posted within a 60Km radius of an arbitrary point within central London. In total, 42,661 tweets were collected within the sample. During the data cleaning phase of the study, all tweets which did not contain geo information were discarded. Of the 42,661 tweets collected in the sample, 2,605 (6.1% of the sample) contained WGS84²⁵ geo information (i.e. a valid latitude and longitude of the point from where the tweet was posted). This low percentage was comparable with a previous study investigating Twitter geodata²⁶, which reported that of a random sample of over 1 million Twitter users, geo information was contained in fewer than 0.42% of all tweets.

The level of precision of the geodata in the sample varied by tweet – from one decimal place to eight decimal places. This meant that in some cases tweets posted from very close locations were considered to have been posted from different locations because of the level of precision. In order to interrogate the tweets to identify possible church locations, the latitude and longitude values

²³ Cooper, 'Unwrapping'.

²⁴ Brierley, *Capital Growth*.

²⁵ The World Geodetic System 1984.

²⁶ Cheng, Caverlee & Lee, 'You are'.

were rounded to four decimal places, giving approximately 11.1 meters of error²⁷. Combined with mobile phone GPS error of up to 8.5 meters²⁸ this meant that tweets were grouped together, and considered within the analysis to have been posted from the same location, if they were actually posted within approximately 20 meters of each other. This proximity has been deemed sufficiently close, for the purpose of this study, given the typical size of a church.

To identify tweets which were posted from the same location, a pivot table was created which included geodata in the rows and the count of tweet ID in the values. In total, 211 geolocations were identified as containing multiple tweets posted during the 23 week period. Where the same geolocation contained multiple tweets, these tweets were manually inspected and qualitatively coded into a 'possible church location' or a 'location of no further interest'. Following this initial qualitative review, 78 geolocations were determined to be possible church locations while 133 geolocations were determined to be of no further interest.

The 78 possible church locations were then plotted on a map in order to identify locations which fell outside of London – a number of such locations were expected at this point in the analysis, as the data collection method used an arbitrary point within London to form the centre of a radius search. At this point, 66 possible church locations were determined to be within London and 12 possible church locations were determined to be outside of London, and were thus discarded from the data set.

The 66 possible church locations within London were then subjected to a second round of qualitative analysis. Each geolocation was considered in turn, and the content of each tweet was inspected to determine whether or not the name of the possible church could be identified. The geolocation was then reviewed on a map and using Google Street View²⁹ in order to identify a likely church location and verify the identified possible church locations. Where this did not yield results, web-based searching was conducted in order to uncover any details which could help confirm the presence, or not, of a church at the possible church locations. During this round of analysis, 24 possible churches were removed from the data set for a number of reasons, including – failing to identify a church, insufficient information to suggest the presence of a church and identifying duplicate locations for the same church. In total 42 of the possible church locations were confirmed as being actual church locations. As such, this answered research question one, by confirming that geolocated Twitter data can be used to identify church locations. The 42 churches identified by this research are shown in table 6.2:

²⁷ The distance between the North Pole and the Equator is 10,000,000 meters. This distance can be divided by 90 degrees, to calculate the approximate distance of each degree – 1,111,11.1m. Dividing this distance again by 10,000, gives 11.1m for four decimal places. As such, a tweet with geodata containing four decimal places has an error of 11.1m.

²⁸ P. A. Zandbergen S. & Barbeau, 'Positional accuracy of assisted GPS data from high sensitivity GPS-enabled mobile phones', *Journal of Navigation* 64, (2011): 381-399.

²⁹ See: <https://www.google.co.uk/streetview/> (last accessed 29 August 2016).

Table 2: The 42 churches identified in this study using geolocated Twitter data.

Church Name	Postcode
All Souls Church	W1B 3DA
Bow church	E3 3AH
Christ Church East Greenwich	SE10 9EQ
Christ Church Radlett	WD7 7JJ
Christ Church Spitalfields	E1 6LY
Christ Church with St Phillip	KT4 8LG
Christ Church, Mermaid Theatre	EC4V 3DB
Dartford Community Church	DA1 2HW
Hillingdon Park Baptist Church	UB10 9LS
Hillsong Bermondsey	SE16 3LP
Hillsong Central London, Theatre Royal, Drury Lane	WC2B 5JF
Hillsong Central London, Dominion Theatre	W1T 7AQ
Holy Trinity Brompton	SW7 1JA
Jesus House for all the nations	NW2 1LT
King's church Walton	KT12 3JB
Macedonian Orthodox Church	W1D 4NQ
Regeneration Church	RM2 6DH
Rivergate Church	IG11 0FJ
Sacred Heart Church	SW19 4LU
Sacred Heart Church	NW6 4PS
St Alfege Church	SE10 9BJ
St Andrew's Church Watford	WD17 4PY
St Bride's church	EC4Y 8AU
St Cedds Church	RM8 2HQ
St George's Catholic Church	HA0 2QE
St James Roman Catholic Church	W1U 3QY
St John the Baptist Greek Orthodox Church	N8 0LY
St John-at-Hackney	E8 1GY
St John's Wood Church	NW8 7PF
St Margaret's	SW1P 3JX
St Martin's Church	HA4 8DG
St Mary Abbots Church	W8 4LA
St Mary the Virgin Church	EN4 8XD
St Mary with All Souls	NW6 4SN
St Matthew's Church Brixton	SW2 1JF
St Pancras Old Church	NW1 1UL
St Paul's Cathedral	EC4M 8AD
St Paul's church	W13 9XW
Temple church	EC4Y 7BB
The Redeemed Christian Church of God, Victory House	SE17 1TJ
United Reformed Church Hounslow	TW3 1UF
Westminster Abbey	SW1P 3PA

In order to identify which, if any, of the 42 churches identified using geolocated Twitter data were missing from the 2012 London Church Census, details of the churches listed in table 6.1 were submitted to the census lead researcher, Dr Peter Brierley, for comparison with the list of churches included. In total, nine of the churches identified in this study were found not to have been included in the 2012 London Church Census. As such, this answered research question two, by confirming that geolocated Twitter data can be used to uncover previously hidden churches. Table 6.3 details the nine previously excluded churches which were uncovered using geolocated Twitter data.

Table 3: The nine previously excluded churches identified in this study using geolocated Twitter data.

Church Name	Postcode
Bow church	E3 3AH
Dartford Community Church	DA1 2HW
Hillsong Bermondsey	SE16 3LP
Hillsong Central London, Theatre Royal, Drury Lane	WC2B 5JF
King's church Walton	KT12 3JB
Macedonian Orthodox Church	W1D 4NQ
Rivergate Church	IG11 0FJ
St Pancras Old Church	NW1 1UL
The Redeemed Christian Church of God, Victory House	SE17 1TJ

The churches identified within this study, both those included in the previous church census and those not included in the previous census, were plotted onto a map in order to visualise the geographical distribution of the churches across London. This revealed that the churches were broadly spread across the London Boroughs, with a concentration within the central zone one area, as might be expected.

Discussion

The results of this study have demonstrated that it is possible to identify the locations of churches using freely available geolocated Twitter data. More importantly, this study has highlighted that the use of such data to identify the locations of churches can complement traditional survey research into church attendance, by identifying churches which would otherwise not be included in such studies. Importantly, this application has the potential to take researchers one step closer towards reducing the level of under counting which currently takes place in church attendance research, and towards a clearer understanding of whether the church is in a state of net growth or net decline; thus, more directly addressing some of the challenges posed by the secularisation thesis. This chapter does not seek, by highlighting the existence of churches which were not previously included, to undermine the findings of previous studies such as the London Church Census. In a contemporary setting where churches regularly form, merge, move, split and close it is inevitable that some churches will not be identified for inclusion in research studies. While it is desirable to reduce the extent of under counting in church attendance studies, a degree of under counting is preferable to over counting as a result of poor methodology or over estimation as a result of inappropriate extrapolation of non-robust data. As such, it is hoped that the methodological insight identified through the research presented in this chapter can be used in future church attendance studies to enhance data quality and give a clearer understanding of the state of the church.

The method of data collection is a limiting factor of this study. The Twitter API provides a free sample of tweets which match a specified search request, rather than the full population of available tweets, which could be purchased through a third-party data re-seller. In returning results to a search request, the Twitter API provides a time slice of all tweets which meet the search request. The way in which that time slice is generated is not clear, and as such it is not clear how random (in statistical terms) the sample is. For studies seeking to extrapolate findings from a sample study across a wider population, this limitation could be of considerable concern. For a methods-focused study such as this one however, the possible lack of randomness is less concerning. The main aim of this chapter was to consider whether freely available Twitter data could be used to identify churches, and in particular, churches which were not previously included in quantitative church attendance studies. The fact that both of the research questions posed were answered affirmatively using the sample of data available provides confidence that the same result would have been possible using a complete set of tweets. In fact, including further tweets in the analysis may even have increased the number of churches which could be found. As such, while this study has identified that the use of freely available Twitter data can be of value to church attendance researchers, future research may benefit further by testing whether or not the use of a full data set provides any further return on investment.

In order to collect a sample of tweets for use in this study, the Twitter API was used to collect tweets posted in the English language each Sunday over a 23 week period, which contained the word *church*. Tweets were restricted to the English language for ease of analysis in one language. Tweets were selected from Sundays as initial exploration on Twitter identified that the majority of tweets containing the word *church* were posted on Sundays. Tweets were selected if they contained the word *church*, as this simple search term created little noise – including more complex search terms could have inadvertently increased the level of irrelevant content, increasing the difficulty of the qualitative components of the analysis. Each of these decisions ultimately played a limiting factor on the volume and breadth of tweets returned by the Twitter API during the data collection phase of the study. Expanding the search term to make it more comprehensive would likely have enabled tweets to be collected which were posted from a number of London's diverse churches which meet on days of the week other than Sunday, which are attended by non-English speakers and which use a nomenclature other than *church*. As this study was a proof of concept to explore whether or not freely available Twitter data could be used to improve knowledge of church locations, this limitation was not considered to be of sufficient concern to jeopardise the research. However, in light of the findings of this study, future research could benefit from repeating the approach described in this chapter, in order to identify whether an expanded search term could deliver improved results, and if so, how much more effective more comprehensive search terms are than the basic search term described in this chapter. Any identified improvements in the approach described in this chapter could, if used alongside future survey research methods, take the research community closer towards understanding whether or not the church is in a state of net growth or net decline.

Within the study, tweets were grouped together into clusters which were posted from the same location by truncating the latitude and longitude for each tweet to four decimal places, and looking for groups of tweets which contained exact latitude and longitude matches at this level of precision. As described previously in this chapter, this meant that within the analysis, tweets were considered to have been posted from the same location, if they were actually posted within approximately 20 meters of each other. Given that many churches across London have a floorspace such that tweets posted from different parts of the same location could be posted greater than 20 meters apart, the approach taken in this study is likely to have introduced a degree of false-negative error – by failing to identify a number of church locations which could otherwise have been identified had the latitude and longitude values of the tweets been rounded to some decimal place greater than four. This decision was, however, deliberate as it is highly unlikely that church-related tweets posted within 20 meters of each other are from separate

church locations, thus reducing false-positive error that church locations are identified incorrectly. Had less precise latitude and longitude data been analysed this may not have been the case – particularly where churches are very closely located, as is the case for many churches in London. Should future studies seek to repeat the approach taken in this chapter whilst experimenting with the precision of latitude and longitude data, consideration of the introduction of false-positive error will be of importance, in order to ensure that the results – identified church locations – remain valid.

The focus of this study was on testing the use of freely available Twitter data to identify churches in London. London was selected as the location with which to test this methodological approach as good data on the church in London already existed which could complement the findings of this study, and as such, because this study would be able to identify whether or not the use of geolocated Twitter data can be of benefit to traditional offline research approaches. Notwithstanding these reasons for selecting London as the city of focus for this proof of concept study, there is no theoretical reason to believe that this approach could not be repeated with similar levels of success for other cities – both in the UK and overseas. As such, future research is recommended to confirm that this approach is a viable and useful tool for identifying churches in other cities worldwide.

Similarly, Twitter data was selected for use within this study because it is quickly and easily available in a consistent format and because a sample of Twitter data is accessible to researchers free of charge. Future research could build on the findings of this study by testing geotagged data provided from other social media platforms, in order to identify whether or not churches can be identified using different sources of social media data.

Finally, while the research presented in this chapter focused on finding churches using geotagged Twitter data, the method used could be applied to a wide range of research challenges, to enable the identification of locations of other groups, organisations or places of interest. This could enable the academic community to increase the reach of their future studies, by identifying further groups/individuals for inclusion in offline studies, and could also be used to help identify a number of hard to reach populations. The development of the method described in this chapter for such uses could thus make social media research a powerful tool for both the research community and policymakers.

Conclusion

The research presented in this chapter investigated whether or not it was possible to identify church locations using geolocated Twitter data, and if so, whether it was possible to identify any previously hidden churches using geolocated Twitter data. The analysis found that it was possible to identify church locations and the details of 42 identified churches are provided. Furthermore, the analysis confirmed that nine of the 42 identified churches had previously not been included in the London Church Census. There are a number of implications arising from these findings which are discussed in this chapter. One implication is the likely under counting of church attendance – the identification of previously hidden churches indicates that data quality challenges mean that not all churches in London have previously been approached in quantitative church attendance studies, and are thus excluded from measures of church growth. A second implication is the possible improvements which could be made to church attendance data quality through the use of social media data – future studies could benefit from repeating the steps presented in this chapter in order to identify churches for inclusion in quantitative studies, thus reducing any future under counting. Finally, this chapter considered the limitations of the research conducted alongside recommendations for future research to build on the findings of this study.