

Beneficiaries and bill-payers

Charities, funding, and Twitter

Natalie Polack
University of Stirling
npp00002@students.stir.ac.uk

Tom Wallace
University of Stirling
tom.wallace@stir.ac.uk

Vikki Richardson
University of Stirling
vip0001@students.stir.ac.uk

ABSTRACT

This paper investigates factors that may determine a charity's use of, and popularity on, Twitter. Previous research has shown that organisational size has a large effect on charity use of social media, with larger charities making more use, but what other factors can predict this behavior? We use a combination of traditional, web-scraped, and API sourced data, along with regression analysis to answer this question. Our results indicate that providing support to the public and receiving funding from the public both appear to be significant in determining charity use of Twitter. In other words, it is a charity's beneficiaries and bill-payers, which best predict levels of Twitter use.

KEYWORDS

Charities, Twitter, funding, data, regression.

1 INTRODUCTION

This research paper explores factors that predict charity use of Twitter. Twitter is a social networking platform that is commonly used by charities but remains understudied academically (Obar, Zube, & Lampe, 2012). Charities have many uses for Twitter, including advertising charitable campaigns, interacting with their users or supporters, and even communicating directly with other charitable organisations. Twitter has a relatively low cost of use, only requiring staff time. This may present an opportunity for smaller charities, which are often limited by costs. Larger charities however, are more likely to have the capacity to dedicate staff to social media and so are still, likely, at an advantage. Our study attempts to predict various aspects of charity Twitter usage (such as overall use, popularity, and networking activity) based on publically available data on charities features (such as funding source, overall income, who the charity helps, and number of staff). This leads to the following research questions:

RQ1: Is funding source, the extent to which a charity is funded by the government or the public, related to the use of Twitter?

RQ2: Are charities that seek to help the public more popular on Twitter?

RQ3: Does number of staff, rather than size, determine a charity's active use of Twitter?

This study is important because there is a lack of published research on the factors, which relate to successful use of social

media by charities. Twitter has huge potential for charitable organisations and knowing which factors contribute to success or failure on the platform could help implement measures that lead to more of the former and less of the latter.

The next part of this paper gives a brief summary of the data management tasks. Section 2 then describes previous literature and research in this area. Sections 3 and 4 detail the data processing and analysis methods used before the results are discussed in Section 5 and the paper is then concluded.

1.1 Data sources and management tasks

Three data sources were required to undertake this research. To obtain data on the outcome, Twitter use by charities, we downloaded data directly through the Twitter API (2018) for the charities in our sample (at least for those that had an account). For data on charity features, we scraped the UK Charity Commission website (2018) which holds a record for every registered charity in England and Wales (Scotland and Northern Ireland are covered by separate regulators). Given these two data sources, we could sample an arbitrary number of charities, right up to the full population of ~168,000 that are registered (The Charity Commission, 2018). However, the Charity Commission website does not record detailed financial information for charities, which is important for RQ1. Therefore, we make use of data collected by Alcock and Mohan of the Third Sector Research Centre (2017). This dataset is based on surveying a random sample of registered charities and contains detailed financial information, such as funding source. As this dataset is the smallest, it forms our sample and we use it as a 'mask' to collect data from the Charity Commission and Twitter on this subset of charities. The Alcock and Mohan data gives us a sample of 12,150. Each of the three data sources needed to be processed differently and this presents different data management challenges, which are detailed in Section 3.

2 LITERATURE AND RELATED WORK

Twitter use by charities is relatively common. The Lloyds Digital Index found that 44% of the, more than one hundred thousand, charities they surveyed in 2016 were using social media in some form (Lloyds Bank, 2016). Of the various social media services, Twitter is by far the most used platform by charitable organisations (Guo & Saxton, 2014). However, what do charities use Twitter for? A common use case is broadcasting one-to-many messages to followers to convey information or solicit donations (Phethean, Tiropanis, & Harris, 2015; Waters & Jamal, 2011).

Another use of Twitter is networking activity between charities, either because they share a common purpose or for support (Honeycutt & Herring, 2009). Infrastructure organisations, which are charities established to help support other charities, can use Twitter to share links to information, resources, and training (Dayson & Sanderson, 2014).

Overall, the literature, which has been published so far, seems to concur that social media use helps make a charity successful (Lloyds Bank, 2016; McCabe & Phillimore, 2012). Therefore, it is important to understand what factors make a charity an active user of social media and this has largely been neglected by the literature thus far. This provides the project's lacuna.

3 DATA PROCESSING & ANALYSIS METHODS

3.1 The Alcock and Mohan data

The Alcock and Mohan data (2017) was sourced from the UK Data Archive (<http://data-archive.ac.uk/>). It contains detailed financial information for charities in England and Wales collected via a survey. The data is longitudinal and contains information from the financial years 2006/2007 up to 2013/2014. For this project, it was decided to use data from just one financial year. The year selected was 2011/12, as the information for the more recent years was less complete and would have provided a smaller sample.

The original data was imported as a Pandas data frame; Pandas was capable of importing the proprietary .dta format in which the data is sourced. Extraneous years were then dropped from this data frame to leave only data from 2011/2012, which resulted in a sample of 12,150. Charity number was used as the index for this data frame and is used throughout this project as an ID number for linkage as it uniquely identifies each charity.

The financial details in the Alcock and Mohan data are split between a large number of variables and the names used in the raw data are not particularly informative. Therefore, income was renamed as 'Income2011-2012' to distinguish it from up-to-date information collected via a web scrape (see section 3.2). A 'government funding' variable was then created by combing the following individual variables: 'funds government sector', 'funds central government', 'funds local government', 'funds regional government', 'funds EU government', 'funds international government agencies', 'funds foreign governments', 'funds devolved government'. 'General public funding' was contained within a single variable and did not need to be combined, so was simply renamed.

Two new variables were then created to assess the proportion of each charity's funding that came from the government or the general public. This was achieved by dividing each charity's government and public income by its total income.

After creation and renaming of the variables of interest, the data was filtered to create a smaller data frame containing the 12,150 charities and 6 variables: 'Charity number', 'Income2011-2012', 'government funding', 'General public funding', 'Proportional government funding', and 'Proportional general public funding'.

3.2 The Charity Commission data

It was desirable to join the Alcock and Mohan data to data scraped from the Charity Commission website for two reasons. Firstly, the Charity Commission website contains updated income data for 2018, which allows growth measures to be calculated, and secondly, it contains variables not found in the Alcock and Mohan data. The new variables found on the Charity Commission website include 'staff', 'who the charity helps', and 'website'. 'Staff' is a simple integer variable, which is found in the Alcock and Mohan data but is unreliable. The version in the Charity Commission data represents fulltime equivalent employees and is collected from account returns, so it is more trustworthy. 'Who the charity helps' is a categorical variable which charities specify when they register from a dropdown list (it is not free-text). Example categories include 'Children/young people', 'Other charities', and 'The general public/mankind'. We are interested in the latter, which is formatted as a binary variable in the final data. 'Website', where provided, is an URL to the charities own website. This was used to obtain Twitter handles as detailed in Section 3.3. Several other variables were collected during the web scrape, which were not used in this research but could be employed by wider projects. These include 'Number of volunteers', 'Number of trustees', 'Expenditure' and 'Company number' which could be used to link to Companies House data in the future.

The web scrape was performed using the BeautifulSoup (<https://www.crummy.com/software/BeautifulSoup/bs4/doc/>) package in Python. The Charity Commission's website is structured so a unique charity number identifies each record. This made it easy to loop over each charity in the Alcock and Mohan data as this contained the same charity numbers. On each loop, data was collected on all of the above variables for a single charity and these were combined into a single dataset. Error handling was built into the scraper to make it robust to missing data or charities that had been removed. It was also robust to disconnection and blocking, but was run over several hours during an off-peak period to avoid overloading the Charity Commission's servers.

3.3 The Twitter data

Gathering Twitter metrics for the sample involved a two-stage process, first collecting handles for charities via a web scrape, and then gathering metrics for these handles through the Twitter API.

Obtaining Twitter handles for charities

In order to obtain a handle for each charity it was necessary to scrape each charity's website using the Python library BeautifulSoup to parse HTML. The JSON file which resulted from the web scrape in 3.2 contained charities' websites (where listed) and was imported as a Pandas data frame. Each of the

URLs was then looped over with BeautifulSoup attempting to find hyperlinks on each charity's homepage to Twitter.com

These hyperlinks formed a list, which was, in turn, iterated over to find relevant Twitter links for the charity. Some websites contain links to Twitter.com for reasons other than linking to their own Twitter account and this was accounted for. If the link contained keywords such as 'search', 'home', 'archives', 'tweet', 'share', 'intent' or 'Twitter' this was likely to point to something other than the charity's Twitter handle and was dropped from the list. Once a relevant link to the charity Twitter handle was found the iteration was escaped and the Twitter handle extracted from the link using regular expressions. If no relevant Twitter handle was found for a charity, it was given a missing value.

Twitter metrics

Once all the Twitter handles were found, the Python Tweepy library (<http://www.tweepy.org/>) was used to interact with the Twitter API (developer.twitter.com) and extract metrics.

Similarly, to the Twitter handle collection method, a data frame was created and the Twitter Handle list was iterated over and passed to the API. Tweepy handles rate limiting internally which proved fortuitous as the Twitter API penalises users who ignore the limits. Getting the metrics was a two-stage process; for each Twitter handle, Tweepy attempted to identify the handle as being valid on Twitter. Once this was found, metrics could be easily obtained with inbuilt commands. These were returned as Pandas series with missing placeholders added for any failed users or where there were instances of having no Twitter handle for which to check. By the end of the iteration four variables had been created 'Twitter Handle', 'Has Twitter', 'Twitter followers', 'Twitter following', 'Number of tweets in total'.

Table 1. Charity handle ownership

	Has Twitter	Does not have Twitter
n	4,875	7,275
%	40%	60%

As shown in Table 1, we found handles for 40% of our sample. We tested our scraping methods against a manually collected sample of 155 Twitter accounts. This sample was collected by manually searching Twitter and Google for each of the 155 charity's Twitter accounts and is 100% accurate. The 155 accounts were selected arbitrarily with a random number generator (based on Mersenne Twister). We compared the results of our scraper with this list of accounts and found that our method correctly identified 67 of the 155 accounts or 43%. While this figure is somewhat low it should be noted that these are all false negatives and our method did not return any false positives, which would introduce bias. Given we sampled thousands of charities, the relatively low accuracy of the scraper should not be problematic for our analysis. Charities are expected to be missing at random and therefore should not bias the results.

3.4 The Combined data

Once the Alcock and Mohan, Charity Commission, and Twitter data had been collected, managed, and stored as individual datasets (in the JSON format), the final task preceding analysis was to combine these different sets of data into one file. This task was eased by the inclusion of charity number in each set; data was only collected from the Charity Commission and Twitter for charities in the Alcock and Mohan data and charity number was carried through all datasets as a unique identifier. Use of missing markers also means that every variable within each dataset was the same length. This meant that combining the data was simply a matter of merging using charity number. There were no duplicates and no fuzzy matching was necessary.

With the data combined, some final data management tasks were performed before analysis. Firstly, absolute and ratio funding growth variables were created. The former is simply income from the updated Charity Commission data minus the income in the Alcock and Mohan data, for each charity. This results in the change in funding for charities between 2011-2012 and 2018, for those present in both datasets. This will help measure charity growth, decline or stability. The funding ratio variable was created by dividing the income from Charity Commission data by the income from the Alcock and Mohan data. This ratio variable was used to measure the same concept but controls for variations in starting income.

With these new variables created, the final task was to deal with outliers. Funding and Twitter data are both commonly outlier heavy and this could unduly bias our analysis. A function was created which removed data points that were more than 1.5 times outside the interquartile range for each variable. This is a conventional way to treat outliers (Everitt & Howell, 2005). The data removed was set to missing rather than being dropped. Testing revealed that this method of outlier removal was more appropriate than using a standard deviation based method. Note that the two government funding variables were not treated for outliers as they were so skewed that treatment resulted in an unacceptably low number of remaining data points. Therefore, care is taken when interpreting the government funding variables in the results. The effect of the outlier treatment is tabulated below.

Table 2. Outlier treatment

Variable	Number of outliers set to missing	Original data points	Percentage lost
Number of tweets in total	367	4,796	7.7%
Twitter followers	602	4,796	12.6%
Twitter following	342	4,796	7.1%
Income 2011-12	1,540	12,150	12.7%
Income 2018	1,342	10,800	12.4%
Funds general public	2,153	12,150	17.7%

Proportion general public funding	2,008	12,150	16.5%
Staff	1,455	10,800	13.5%
Absolute funding growth	2,222	10,800	20.6%
Ratio funding growth	703	10,800	6.5%

The box plots below show an example of the effects of outlier treatment on the variable 'Funds general public'. This had a relatively high level of loss from the function (17.7%) and still has outliers after treatment, but is much closer to normal and should be suitable for analysis.

Figure 1. Boxplot before outlier treatment

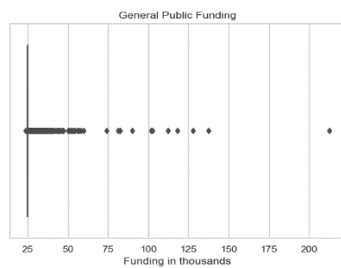
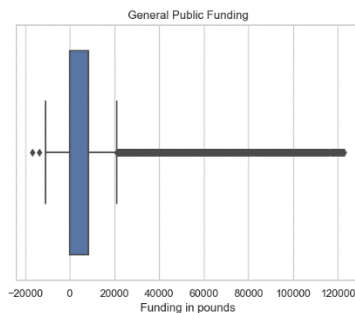


Figure 2. Boxplot after outlier treatment



4 ANALYSIS METHODS

This section briefly details the statistical methods that we applied to our data. Starting with basic descriptive methods and culminating in a discussion of our modelling techniques.

4.1 Univariate methods

The focus of our analysis is regression, therefore we only use brief univariate methods to describe the data before analysis. These methods include statistical summaries for metric variables, which show the distribution, mean, median, standard deviation, and extreme points of the data. Histograms are also used in this endeavour, visually showing the distributions of the metric variables such as income, number of Tweets, etc. Finally, box plots are also employed to demonstrate the spread of data and the effect of outlier removal. For our categorical data, we use one-way tables.

4.2 Bivariate methods

Bivariate methods fill a similar role to the descriptives; they describe relationships between key variables before they are fully modelled with controls. These methods are not particularly robust, as they lack controls, and are therefore used as primers for the modelling rather than fully inferential components of analysis. As most of our data is metric, we mostly use Pearson's correlations for this task with pairwise deletion of missing values.

4.3 Multivariate methods

The weight of our analysis rests on regression modeling. This is because models not only provide more detailed output than simpler methods, but also allow for more than one independent variable to be tested against the dependent simultaneously, with the effects controlling for each other. This ability to control for other effects is critical to one of our research questions, which seeks to investigate the effect of staff numbers net of income.

We use two types of model, depending on the format of the dependent variable. For metric outcomes, we employ ordinary least squares, which is the most conventional regression model, predicting the conditional mean. For binary categorical outcomes, we use logit, which predicts the probability of being in the affirmative category of the outcome rather than variation within a metric scale. These two types of model have similar outputs and are interpreted similarly. Where they do differ, we note this in the results.

5 RESULTS & DISCUSSION

5.1 Question 1: How is source of funding related to charity use of Twitter?

The aim of this research question was to discover if government funding or general public funding had an influence on the likelihood of the charity having and using Twitter. Therefore only those charities that were government or public funded were retained for analysis. It seemed plausible that charities relying more heavily on public funding would be more inclined to have and use Twitter as a way of raising funds and this question will assess the extent to which this is true.

Having a Twitter handle

The table of the summary statistics for government funding and general public funding in this subset of data is displayed below.

Table 3. Summary statistics for funding

	Funds general public	Govt funding	Prop general public funding	Prop govt funding
n	4,087	4,120	3,381	4,120
Mean	27,990	14,726	0.03	0

Median	14,880	0	0.02	0
Std Dev	31,506	346,807	0.04	0.04
Min	0	0	0	0
Max	122,848	19,190,000	0.15	0.90

The funding variables are still quite positively skewed, therefore using proportional forms of these variables was deemed most appropriate. We note that 98% of our sample has some general public funding while only 5% has some government funding. Therefore, this data set is biased towards charities funded by the general public.

Table 4. Having Twitter

Has Twitter	Government funding	Public funding
No	48%	60%
Yes	52%	40%

Investigating the proportions of each funding category to measure having Twitter shows, at a glance, that government funded charities are more like to have Twitter. However, regression results will need to be consulted to fully understand if this is true.

Model 1. Charity funding and handle ownership

Dependent: Has Twitter	Coefficient	Std Error	P-value
Proportion of government funding	-0.35	1.07	0.74
Proportion of general public funding	0.95	1.05	0.36
Income 2018	<0.001	<0.001	<0.001
Constant	-0.52	0.06	<0.001

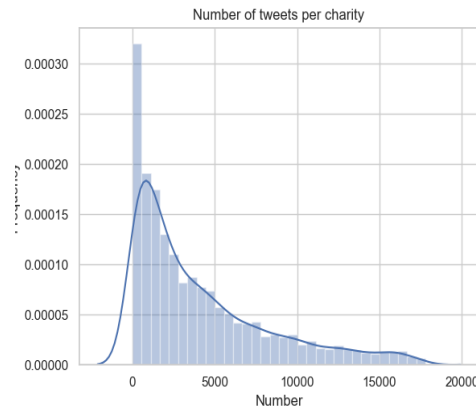
Pseudo R-Squared = 0.032

The model analysed proportions of funding using income as a size control with the outcome variable being the binary 'Has Twitter'. Whilst there is a positive effect of having general public funding, it is not statistically significant and likewise with government funding whilst there is a negative effect it is not statistically significant. We can therefore conclude little from this model as the relationships are not substantiated.

Use of Twitter

Charity interaction with Twitter can also be measured in terms of use, rather than simply having an account. 'Use of Twitter' is the number of tweets the charity has published in total. This is a metric variable, which acts as a proxy for use. Number of tweets is shown in Figure 3.

Figure 3. Histogram of number of tweets



For this question only the charities that were public or government funded and had a Twitter handle were used. All of the variables are positively skewed indicating that most charities have a low proportion of public funding, an even lower proportion of government funding and most charities do not have a high number of tweets; there are a small number of extreme cases in each variable.

Inspecting the correlation of funding variables shows no notable relationship between government funding and number of tweets (this table is not shown for brevity) with only a 4% correlation between public funding and number of tweets.

Model 2. Charity funding and number of Tweets

Dependent: Number of Tweets in total	Coefficient	Std Error	P-value
Proportion of government funding	3,824	2,964	0.19
Proportion of general public funding	8,145	2,953	0.01
Income 2018	<0.001	<0.001	< 0.001
Constant	2,453	193	< 0.001

R-squared = 0.051

The results of model 2 show that general public funding has a significant positive effect on total number of tweets. A charity with a higher proportion of public funding will make, on average, more tweets than one with less income generated from the public. Based on these results there we cannot make any claim about government funding as it relates to activity on Twitter as this result is insignificant. Income is significant and positive, but very small (likely due to the scale of the variable). However, this is a control for charity size and does not need to be interpreted.

5.2 Question 2: Are charities which seek to help the public more popular on Twitter?

This question investigates the link between public-facing charities and popularity on Twitter. We could assume that charities which seek to help the general public will be more heavily incentivised to engage with Twitter to interact with the public. This would result in them being more popular on Twitter (measured by follower count) than non-public-facing charities but to what extent does the data substantiate this assumption?

Univariate descriptives

Table 5. Support for general public.

	Helps the general public	Does not help general public
n	5,528	5,272
%	51.2	48.8

The table above summarises the binary ‘Helps the general public’ variable, which records if a charity seeks to help the general public or a more focused group (such as other charities, children, animals, etc.). Charities such as educational trusts are also usually excluded from this group. Roughly half of the sample is in each category, which should provide good variation for modelling. Summary statistics for these two groups are tabulated below.

Table 6. Summary of variables for nonpublic-facing charities

Does not help general public	Income 2018	Proportion of general public funding	Twitter followers
n	4,612	5,747	1,976
Mean	1,407,509	0.014	2,735
Std	1,687,574	0.029	3,358
Median	758,100	0	1,508
Min	0	-0.046	0
Max	7,400,000	0.151	17,324

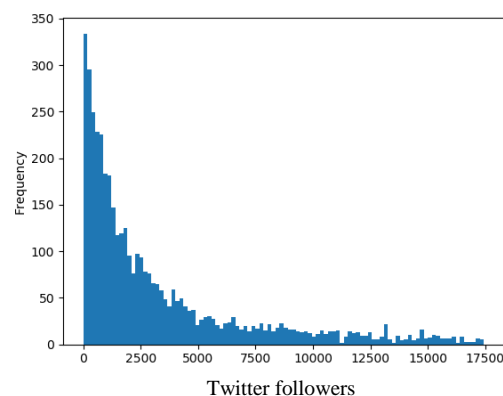
Table 7. Summary of variables for public-facing charities

Helps general public	Income 2018	Proportion of general public funding	Twitter followers
n	4,846	5,544	2,218
Mean	1,297,389	0.018	3,788
Std	1,622,773	0.034	4,118
Median	664,300	0	2,134
Min	0	-0.024	0
Max	7,400,000	0.151	17,446

The two summary tables above are split by the binary ‘Helps the general public’ variable to show how these groups differ in terms of funding and popularity on Twitter. As shown, charities which help the general public appear to have slightly lower incomes, but gain more of their income from public funding. Most importantly, they appear to have more followers on Twitter, but this will be fully explored in the modelling as none of the means in these

tables are tested to ensure they are statistically differentiable. This is performed in the modelling.

Figure 4. Histogram of Twitter followers



This histogram shows the distribution of the dependent variable, Twitter followers, for all charities (after it had been treated for outliers). The distribution appears half-normal, is not overly afflicted with outliers, and should be suitable for modelling. Followers is a good proxy for popularity because it records how many other accounts have chosen to actively subscribe to a given charity’s content on Twitter.

Multivariate modelling

Model 3. Helps the general public

Dependent:	Coef.	Std error	P-value
Twitter followers			
Helps general public (binary)	1,193	121	<0.001
Income 2018	<0.001	<0.001	<0.001
Constant	1,632	110	<0.001

R-squared = 0.064

This is an ordinary least squares model, which predicts number of Twitter followers (popularity) based on whether the charity helps the general public. Income is included as a control for size, which usually has a large effect on Twitter popularity (larger charities tend to be more popular on Twitter). Both the primary independent and the control are significant in this model. The primary measure is binary so the result of 1,193 means that charities which help the general public, on average, have 1,193 more Twitter followers than those who do not – controlling for size. This suggests public facing charities are more popular on Twitter and positively affirms research question 2. However, the R-squared for this model is small which suggests there are many other factors (or simply random variation) which affect charity popularity on Twitter.

5.3 Question 3: Does number of staff, rather than size, determine a charity's active use of Twitter?

This final question attempts to determine to what extent staffing numbers are predictive of activity on Twitter. Previous research by Wallace and Rutherford (2018) found that size, as measured by income, is heavily indicative of a charity's level of social media activity. However, as Twitter is free to use, it is unlikely that a financial advantage would fully explain this effect. This question, therefore, uses information on staff numbers to attempt to explain this effect.

Univariate descriptives

In order to determine whether the number of staff employed by a charity, rather than the size of the charity, plays a role in a charity's successful use of Twitter, three variables were selected from the final data set: 'Staff', 'Twitter following', and 'Income2018'. 'Twitter following' is the number of accounts the charity follows and is used as a measure of active use of Twitter, as some charities may choose not to Tweet very often but may still follow other accounts which shows a level of use.

Table 8: Summary statistics for selected variables

	Staff	Twitter following	Income2018
n	10,695	4,454	9,458
Mean	14.73	1,000.06	1,351,087
Standard deviation	23.89	885.69	1,655,516
Minimum	0	0	0
25%	0	289	138,250
50%	0	758.5	706,700
75%	22	1,505.5	1,900,000

Bivariate Correlation

The correlation between the number of 'Staff' and the number 'Twitter following' was calculated and the result was positive but weak at 0.07. This suggests there may be some form of association but this needs to be assessed in a more robust modelling framework.

Multivariate modelling

A linear regression was specified using 'Staff' to predict 'Twitter following' and the results are shown in Model 4.

Model 4: Results of the Linear Regression Analysis

Dependent:	Coef.	Std error	P-value
Twitter following			
Staff	2.195	0.576	<0.001
Constant	920	20	<0.001

R-squared = 0.004

'Staff' has a strong, positive coefficient, and is highly significant. This suggests that the more staff a charity has, the greater their use of Twitter. However, previous research has shown that the size of a charity is heavily predictive of Twitter use (Wallace & Rutherford, 2018). As size also affects number of staff, it is possible that size is being expressed through staff in the results of Model 4. To disentangle these effects a further model was specified which included 'Income 2018' as a control for charity size.

Model 5: Results of the Linear Regression Analysis

Dependent:	Coef.	Std error	P-value
Twitter following			
Staff	-0.992	0.756	0.189
Income 2018	<0.001	<0.001	<0.001
Constant	867	21	<0.001

R-squared = 0.017

In this model, the coefficient for 'staff' is negative but this is insignificant so we cannot claim to have found any evidence of an association between staff numbers and Twitter popularity. Income has a very small positive and significant result, which suggests that larger charities are more active on Twitter. However, this variable was a control and the main effect is insignificant so we cannot claim to have found any evidence that staffing has an impact on Twitter use by charities beyond its association with size.

6 CONCLUSIONS

6.1 Summary

This research paper attempted to determine which factors, beyond size, have an impact on charity use of, and popularity on, Twitter. The influence of funding type on the likelihood of a charity having and using Twitter was addressed first. Government funded charities were found to be more likely to have a Twitter account. However, it was general public funding that was found to have a significant positive effect on the total number of tweets a charity will make. This measure reflected their activity on Twitter. Both helping the public and the size of a charity were found to be significant factors when predicting the popularity of a charity on Twitter. On average, those charities that are more public facing have more Twitter followers than those who do not. Finally, the effect of the number of staff on a charity's active use of Twitter was considered. Using the number of staff to predict the number of accounts that a charity follows on Twitter initially suggested that the number of staff was a significant variable. However, when controlling for size using income, no evidence was found that number of staff was a significant predictor of Twitter use. Overall, we conclude that it is a charity's beneficiaries and bill-payers, which seems to determine the level of its use of Twitter.

6.2 Limitations

Despite the apparent strengths of this research, our methodology has some necessary limitations. The Alcock and Mohan data is several years out of date and we were forced to compare proportions of funding from 2011-12 to Twitter use in 2018. There is no more up-to-date source of data with such detailed financial indicators. The Twitter scraper used charity websites to detect Twitter handles but there are charities who do not maintain their own websites yet have Twitter handles. Charities are also not obliged to give their website to the charity commission. Twitter metrics for these charities will have been omitted from our dataset. However, these missing cases are false negatives, and as long as they are missing at random should not colour the results.

6.3 Future work

This was a relatively modest piece of research and could be expanded or built upon in a plethora of ways. Some of the most promising next steps relate to the information collected by the scraper which we did not use. This includes company number (many charities are companies as well as registered charities) which could be used to link to companies house data. The Charity Commission website also contains Trustee information, which could be analysed. Twitter also provides a rich opportunity for network analysis, which we did not undertake. Rather than simply looking at numbers of tweets of followers, network analysis would map connections between charities and the public and allow us to quantify these relationships. Finally, there is a large amount of text relating to charities available for research. This includes the text on charity websites, in the tweets they make and in their Twitter profiles. This complex form of data has long been too difficult for traditional statistical methods to handle, but the advance of machine learning makes this data available for research. We envisage a text classifier which uses text content from charity Twitter accounts to classify use of Twitter; are charities engaging with their users or simply broadcasting, and what features predict differences in this Tweeting behaviour?

REFERENCES

- Alcock, P., & Mohan, J. (2017). *Third sector research centre research data collection*. Colchester: ESRC.
- Dayson, C., & Sanderson, E. (2014). Working paper 127: Building capabilities in the voluntary sector: A review of the market.
- Everitt, B. S., & Howell, D. C. (Eds.). (2005). *Encyclopedia of statistics in behavioral science*. New York: John Wiley & Sons, Ltd.
- Guo, C. & Saxton, G. D. (2014). Tweeting social change: How social media are changing nonprofit advocacy. *Nonprofit and Voluntary Sector Quarterly*, 43(1), 57-79.
- Honeycutt, C., & Herring, S. C. (2009). Beyond microblogging: Conversation and collaboration via twitter. Paper presented at the *IEEE System Sciences, 2009. HICSS'09. 42nd Hawaii International Conference On*, 1-10.
- Lloyds Bank. (2016). *UK business digital index 2016*. Lloyds Bank.
- McCabe, A., & Phillimore, J. (2012). *Seeing and doing: Learning, resources and social networks below the radar*. Third Sector Research Centre.
- Obar, J. A., Zube, P. & Lampe, C. (2012). Advocacy 2.0: An analysis of how advocacy groups in the united states perceive and use social media as tools for facilitating civic engagement and collective action. *Journal of Information Policy*, 2, 1-25.
- Phethean, C., Tiropanis, T., & Harris, L. (2015). Engaging with charities on social media: Comparing interaction on facebook and twitter. Paper presented at the *Springer International Conference on Internet Science*, 15-29.
- The Charity Commission. (2018). Registered charities in england and wales.
- Twitter. (2018). Twitter API reference.
- Wallace, T., & Rutherford, A. (2018). *The big bird gets the worm? how size influences social networking by charitable organizations*. Unpublished manuscript.
- Waters, R. D. & Jamal, J. Y. (2011). Tweet, tweet, tweet: A content analysis of nonprofit organizations' twitter updates. *Public Relations Review*, 37(3), 321-324.