**Birkbeck, University of London**

**Department of Computer Science and Information Systems**

**Who is following the Government Cats?**

**A Community analysis of the United Kingdom Government Cats’**

**Twitter pages**

**Project Proposal**

**Pravin Jeyaraj**

**MSc Data Science**

**April 2019**

**This proposal is substantially the result of my own work, expressed in my own words, except where explicitly indicated in the text. I give my permission for it to be submitted to the JISC Plagiarism Detection Service.**

# Introduction

Watching cats, especially when they are made to appear to do human-like things, has become known as one of the most popular pastimes on the internet. More cat pictures are shared online than selfies.[[1]](#footnote-1) In an online survey completed by almost 7,000 Facebook and Twitter users, Jessica Myrick found that the average frequency for looking at internet cats was two to three times a week to daily.[[2]](#footnote-2) The first cat video was uploaded to Youtube by its founder, Steve Chen. Now there are over two million cat videos on YouTube being viewed 26 billion times.

The trend for “cat content” predates and is not exclusive the internet. Edith Podhovnik has documented the strong presence of cats in historical accounts, literature, art and media.[[3]](#footnote-3) The government departments’ Twitter pages for their own departmental cats sits well within this context. Media interest in the government’s own cats fits into this context. From time to time, the British media cover of politics from the point of view of the cats. Indeed, just as in real-life politics, there seem to be interest in the competition and rivalry between the different cats.[[4]](#footnote-4) There are currently a reported five cats, owned by the civil servants in five government departments, as indicated in the table below, although the National Archives records show that different government departments have been using cats to catch mice since the 19th century. Reports suggest that there has been a Downing Street or Treasury Cat since the reign of Henry VIII.[[5]](#footnote-5) Public followings of the governmental cats can be traced back to at least 1947, with the Home Office’s Peter III and official records for the upkeep of the cats began in 1936.

The current government cats are[[6]](#footnote-6):

|  |  |
| --- | --- |
| Name of Cat | Government Department |
| Larry | 10 Downing Street |
| Palmerston | The Foreign Office |
| Gladstone | The Treasury |
| Ossie and Evie | The Cabinet Office |

The aim of this study is to develop an understanding of the communities of the different government cats. This understanding can be broken down into the following aspects:

* An initial analysis of each of the government cats’ twitter pages, with regard to descriptive and graphical statistics;
* A distributional analysis of each of the government cats’ twitter followers, assuming that the cats are at the centre of their own respective local networks, with a view to identifying centres of the global network.

As can be seen in the review of available literature, research has been undertaken into the wider internet cats’ phenomenon as well as the use of Twitter in various contexts, including politics, but very little research has been done around the subject of government cats’. The current generation of government cats’ are also the first of the social media era, which provides an opportunity to learn more firsthand information about the public followings using information provided by the followers.

.

# Background

## Internet cats

Podhovnik undertook a blog-based, research project into the plethora of internet cat content. First, she used the English, Russian and German Yahoo sites to collect one month’s worth of news stories, with the keyword “cat” or German or Russian equivalent and then carried out a qualitative data analysis to categorise them. She also collected the hashtags used to share cat content on Facebook and Instagram. The aim of this study was purely to indicate that there is a lot of cat content on the internet and that the concept of cat content is not unique to the internet.

The adoption of cat content in the media and on social media platforms, particularly in relation to politics, has led to creation of a new word. According to Podhovnik, ‘the term “meowlogism” refers to the word formations which, in a cat inspired word play, have been given a purified, or catified, twist, such as “paws” (pause) and “top cat”…”. She has examined the use of meowlogisms in the #blackcat community on Instagram. For this qualitiative study, the data collection method appears to be manually cutting and pasting posts rather than scraping using data science tools. A selection of tweets from Palmerston’s @Diplomog twitter page reveals a usage of meowlogisms:





In Myrick’s study,[[7]](#footnote-7) it was found that internet cat videos did offer a positive effect on emotions in terms of happiness and stress relief. Respondents also showed an existing affinity towards cats, although did not extend towards having one as a pet. At the same time, they also provided an avenue for procrastination, and thus could create feelings of guilt. The aim of the study was explore whether empirical evidence of accessing online cat content backed up existing psychological theories regarding mood management. The potential for a mood management effect of the government cats’ has been reported in the media:

“‘At a moment of political tension or political conflict, as we have in the U.K. right now at the moment, particularly post-Brexit, this softer side of politics, softer form of communications is quite attractive to government and the civil service,’ says Nick Anstead, an expert in political communications at the London School of Economics. ‘It humanises government at a point where government seems remote and seems to be something that people are alienated from.’”[[8]](#footnote-8)

It is worth noting, however, that only one of the cats, Palmerston, has an official twitter page. All of the other cats’ Twitter pages are described as parody or unofficial pages. Whilst the official Palmerston page is overwhelmingly positive, the unofficial pages appear to be ways of anonymous civil servants letting off steam.

Myrick has identified that her own study is limited because the vast majority of respondents were female and Caucasian and it depended on their subjective opinions rather than objective observations. It also depended on people actually responding to the survey. Neverthleless, she proposed that further study could involve sentiment analysis.

Lori Kogan et al found that showing funny or cute animal videos from Youtube in the middle of a veterinary pharmacology lecture positively affected the students’ mood as well as their interest and self-reported understanding of the study material.[[9]](#footnote-9) This research was consisted of a predominantly female student cohort (20 men, 113 women) and the videos were selected by the researchers rather than the students themselves.

## Government cats

Peter Wells, an open data policy expert and consultant for the Open Data Institute, has actually created a register of the UK government cats – both “active” and “inactive”. He has identified 23 cats in total.[[10]](#footnote-10) His aim was primarily to demonstrate the capabilities of open data and various open data tools and the data is available on Github. Whilst criticisms can be made of the quality of the data,[[11]](#footnote-11) they nevertheless offer a good starting point for data scientific research into the UK government cats. The difference between Wells’ register and this study is that Wells was specifically focusing on producing a list of government cats rather than looking into their social media identities or their Twitter communities.

## The topology of Twitter

When Twitter was set up in 2006, it described itself as a microblogging platform. A blog, short for weblog, is like an online diary: it allowed writers to regularly publish pieces of text that are date- and time-stamped. A microblog on the Twitter platform was essentially a very short blog piece of maximum of 140 characters. (This character length has since been extended, but the basic premise is still the same.) On Twitter, a miniblog can be a standalone piece made up of textual and/or non-textual content (tweet).

Twitter is, in general, regarded as a public platform in that anyone, regardless of whether they are a Twitter user or not, can see your tweets on your Twitter page. Similarly, you do not have to be a Twitter user to be able to see users’ Twitter pages. However, only Twitter users can tweet and engage with other user’s tweets by:

* forwarding someone else’s tweet to other users (a retweet)
* a retweet with your own comment
* a tweet that refers to another user or responds to another user’s tweet (a mention)
* Liking a tweet

You can specifically request to be kept updated of particular users’ tweets, retweets and likes, without having to keep checking their Twitter page, by becoming a “follower”. Their tweets and retweets will then appear in your timeline (essentially a newsfeed). Similarly, your tweets and retweets will appear in the timelines of your followers. In general, you can follow anyone you want and anyone can follow you, without any requirement to “follow back”. As a result, relationships can be unidirectional, bidirectional or non-existent.

There is some control a user can have over who follows or engages with them:

* A user can make their tweets private so that they are only viewable to those whom they give permission.
* A user can mute someone so that the latter’s tweets are do not appear in the user’s timeline, like letting someone make all the noise they want without disturbing the user.
* The most extreme form of control is that user can block other users – neither can see each other’s tweets and blocked users are unable to follow those who have blocked them.

## Community analysis on Twitter

Acer and Deguchi found that younger users used Twitter for real-time communication, while older users used it for information gathering and tweeted less often.[[12]](#footnote-12) Furthermore, in a comparison between users from a more individualistic society (USA) and users from a more collectivistic society (Japan), they found that American users’ tweets showed more individualism when personal uniqueness was taken into account. But, when personal uniqueness was excluded, then Japanese users’ tweets showed more individualism as they placed greater value on personal privacy.[[13]](#footnote-13) There may have been a direct link between value on personal privacy on social media and amount of collectivism in society at large. The methodology employed by Acer and Deguchi comprised:

* An exploratory search for categorical keywords using Twitter’s search engine;
* Identifying and selecting the first 100 Japanese users and first 100 American users in search results, excluding spammers, bots, marketers, businesses, organisations;
* Excluding users who had less than 10 friends or followers, had not tweeted in seven days and had not posted at least 500 messages;
* Establishing a male/female ratio of 50:50, based on user name, profile and picture;
* The retrieval of the last 20 tweets by each user (the maximum allowed by Tweepy), excluding direct replies;
* Coding the data into 7 categories, 2 sub-categories and 5 genres.

In a comparative study of five destination marketing project, Sevin found that these type of projects use Twitter to mostly share news about events and they focus on distributing information rather than developing interpersonal relationships.[[14]](#footnote-14) The data measured comprised:

* Whether tweets contained outside links (and thus conveying information from external sources)
* Whether tweets mentioned other users and whether they were retweets
* Whether tweets contained hashtags
* Whether tweets were tailored towards particular users (through mentions)
* The content of the tweets

The possibility of different uses for Twitter raises the question of a blurring of the boundaries between different uses. How do we know where an interaction between users is simply about receiving and passing on information – as in retweeting news media – and where the users involved form an actual community? Du, Xie, Li, Zhu and Lim argued that it is possible to develop a model of a particular user’s overall interests by not only looking at the news media and celebrities they follow but also those whom the user’s offline friends follow.[[15]](#footnote-15) This, of course assumes that users follow offline friends and contacts. In my view, one of the advantages of Twitter is being able to connect to people who you would not meet offline. In Du et al’s research, Twitter users for study was extracted by crawling a Google tool called Freebase (no longer active) as well as search for Twitter users using the search engine Bing’s API. This method provided 100 million Twitter accounts. Real-time data, such as network structure and recent tweets, about each Twitter user was then obtained using a Twitter API. It is not clear from the paper how this was done, given that Twitter only allows accessing their API in blocks of 15 searches. A user’s offline community was then determined using the following closeness algorithm:

1. Compute the closeness score between u and all other users as well as v, where v represents a connection to u almost as weak as any off-line real-life friend should be;
2. Rank the users in decreasing order of closing score, with those ranked before v being an offline community w.
3. Compute the closeness score between w and every other user
4. Any user coming above v are added to w
5. Repeat until no more additions to w occur

Komorowski, Huu and Deligiannis have also examined the shift of offline communities onto Twitter, in this case within the media industry.[[16]](#footnote-16) They have described this sort of community as a “community of practice”, as opposed to one’s group of family and friends. According to the literature, the common characteristics of a community of practices are a shared domain of interest, to which they are committed, a lively community (i.e. a degree of activity and interaction) and the members are practitioners such that the community becomes a resource (196). Komorowski et al’s research focused on a specific monthly meetup of videogame developers in Brussels. They chose to use a combination of quantitative and qualitive methods. First, media professionals in Brussels were interviewed through a survey to identify which “communities of practice” they participated – 19 were identified from 423 respondants. Each community of practice was then analysed using desk research to see whether they meet the requirements – the four that were identified were then then observed at events. Data was then gathered through crawling Twitter, using a Python script and the REST APIs and stored in well-formatted text files. The data comprised: account screen name, account description, language, location, the number of tweets and likes as well as the tweets and retweets themselves of the four community of practice accounts and the followers of the accounts. Statistical analysis and word frequency analysis was undertaken using Tableau and Nvivo. A social network analysis was carried out, looking at the interaction between users through following, liking, retweeting and mentioning and measuring the correlation between the different variables. Text data analysis to identify particular topics of tweets was done using Latent Dirichlet Allocation.

But as I indicated earlier, a community on Twitter does not necessarily have to have some real world or off world basis. Ch’ng (2015) examined the bottom-up formation of a community through the use of particular hashtags (or keyword or phrase). The object of Ch’ng’s study was a community of teenage girls who had developed crushes on the younger Boston bomber. how many times they tweeted using that hashtag and how many of those tweets have been replies to someone else. Ch’ng’s study is useful because it adopted a particular testable definition of community as consisting of four attributes – membership, influence, reinforcement and shared emotion connections.[[17]](#footnote-17) Ch’ng’s methodology is outlined below.

* Streaming software to map Twitter users (egos) and tweets as nodes, with edges representing links between users and tweets. The edges were directional to indicate the flow of information.
* Only tweets containing particular hashtags were included in the study.
* The mapping was undertaken over series of a series of 15 days, resulting in 60 longitudinal datasets.

# Project Details

The primary data source will be the Twitter pages of the current government cats’ mentioned above.

|  |  |  |
| --- | --- | --- |
| Official | | |
| Palmerston | @DiploMog | 81.4K |
| Parody/Unofficial | | |
| Evie the Cat | @HMCabinetCat | 35.8K |
| Larry the Cat | @Number10cat | 245K |
| Palmerston the Cat | @PalmerstonFOCat | 35.6K |
| Palmerston the FCO Cat | @PalmerstonCat | 9,509 |
| Gladstone | @TreasuryMog | 41.1K |
| Gladstone | @HMTreasuryCat | 19.5K |

In the first stage, I am interested in gathering data for seven simple datasets for each of the above twitter pages, containing a list of each pages’ respective followers and whether each follower is being followed also by the respective government cat. From this, I seek to create a 7 x n matrix, where n is the number of followers for each cat. Each matrix cell will be filled with either 0, 1 or 2, indicating the number of connections between the follower and the cat. 0 will mean a user is neither following nor being followed by a particular cat, 1 will mean that it is following but not being followed and 2 will mean that it is both following and being followed. This matrix will then be used to create a visualisation of the network, differentiating the friends (those followers whom the government cat is also following) and those who are just followers as well as identify the likely centres of the global network (those who are following more than one cat).

In the second stage, data will be gathered on the locations of the followers, as specified by each follower in their profiles. The main limitation is that not all Twitter users state their locations. This may mean that, depending on the ratio of known to unknown locations, the analysis is undertaken either on an incomplete dataset or on a complete dataset with assumptions made about the unknown locations. In the latter case, a simple assumption could be to take the unknown locations as being in the United Kingdom – it is reasonable to believe that the majority of the followers will be from the U.K given that the twitter pages represent an aspect of a UK government department. An alternative assumption could be based on the scanning of the follower’s tweets or of the follower’s followers – users are most likely to be followed by other users from the same country. The objective of this second stage is combine the location dataset with the matrix created in the first stage to plot the network from the first stage on a map.

The third stage would be to create a dataset of the profiles of the followers of each cat in order to undertake a textual analysis. This would help to identify common occurrences of keywords in order develop some kind of follower profile.

Time allowing, the fourth stage would be to create a dataset of the tweets of each government cat in order to identify tweets which are replies to other users or retweets of other user’s tweets. From this, the aim is to compare the replies and retweets with the results in stage 1 and 2.

Twitter has provided a plethora of Application Programming Interfaces (APIs) through which Twitter data can be accessed using various programming languages and tools. More details are available at <https://developer.twitter.com/en/docs.html>. A number of Python packages have been built for interfacing with the Twitter API; these are listed at <https://developer.twitter.com/en/docs/developer-utilities/twitter-libraries.html>. Prior to data gathering, it would be useful to undertake comparison testing of the different packages in order to determine how choice of package affects performance and results, given the rate limits deployed by Twitter.

In light of the above, the interim plan of works for the project is as follows:

|  |  |  |  |
| --- | --- | --- | --- |
| Stage | Action | Duration | Deadline |
| 0 | Testing of python twitter libraries |  | 30 June 2019 |
| 1 | Data gathering and analysis of follower lists |  | 14 June 20 g19 |
| 2 | Data gathering and analysis of locations |  | 28 July 2019 |
| 3 | Data gathering and analysis of profiles |  | 11 August 2019 |
| 4 | Data gathering and analysis of tweets |  | 25 August 2019 |
| 5 | Writing up of report | 26 August – 5 September |  |
| 6 | Submission of Report |  | 10 September 2019 |

# Bibliography

Acer, Adam; Deguchi, Ayaka; ‘Culture and Social Media Usage: Analysis of Japanese Twitter Users’, *International Journal of Electronic Commerce Studies*, 2013, 4, 1, 21-32

Ch’ng, Eugene, ‘The Bottom-up formation and maintenance of a Twitter community: Analysis of the #FreeJahar Twitter Community’, *Industrial Management & Data Systems*, 2015, 115, 4, 612-624

Davies, Caroline, ‘More questions over how No 10 handled the kitty’, *The Telegraph*, 24 November 1997, available at: *The Internet Archive*, <https://web.archive.org/web/20071205223254/http://www.telegraph.co.uk/htmlContent.jhtml?html=%2Farchive%2F1997%2F11%2F24%2Fnmog124.html>

Du, Juan; Xie, Wei; Li, Cheng; Zhu, Feida; Lim, Ee Peng; ‘TwiCube: A real-time Twitter online community analysis tool’, *Database Systems for Advanced Applications: 18th International Conference*, Proceedings, 2013, Part II, 7826, 2, 458-462

Horton, Helena, ‘There is a special door for cats in Downing Street, Nick Clegg claims’, *The Telegraph*, 27 September 2016, <https://www.telegraph.co.uk/news/2016/09/27/there-is-a-special-door-for-cats-in-downing-street-nick-clegg-co/>

Kogan, Lori R; Hellyer, Peter W; Clapp, Tod R; Suchman, Erica; Mclean, Jennifer; Schoenfeld-Tacher, Regina, ‘Use of Short Animal-Themed Videos to Enhance Veterinary Students’ Mood, Attention, and Understanding of Pharmacology Lectures’, *Journal of Veterinary Medical Education*, 2018, 45, 2, pp188-194, <https://jvme.utpjournals.press/doi/abs/10.3138/jvme.1016-162r> [Last accessed: 02/02/2019]

Komorowski, Marlen; Huu, Tien Do; Deligiannis, Nikos; ‘Twitter data analysis for studying communities of practice in the media industry’, *Telematics and Informatics*, 2018, 35, 1, 195-212

Myrick, Jessica Gall, ‘Emotion, regulation and procrastination, and watching cat videos online: Who watches Internet cats, why and to what effect?’, *Computers in Human Behaviour*,2015, 52, 168-175, <https://www.sciencedirect.com/science/article/pii/S0747563215004343> [Last accessed: 14 January 2019]

Podhovnik, Edith, ‘The Purrification of English: Meowlogisms in Online Communities’, *English Today*, 2018, 1-15

Podhovnik, Edith, ‘The Meow Factor – An investigation of cat content in today’s media’, *Arts and Humanities Conference*, Proceedings, Venice, 2016, 127-139, <https://iises.net/proceedings/arts-humanities-conference-venice/table-of-content/detail?cid=38&iid=013&rid=6257> [Last accessed: 14 January 2019]

Porter, Harriet, ‘Why cool cats rule the internet?’, *The Telegraph*, 1 July 2016, <https://www.telegraph.co.uk/pets/essentials/why-cool-cats-rule-the-internet/>

Scott, Lauren, ‘A softer side of government: How Larry the Cat became a purr-fect political companion on Downing Street’, *CBC News*, 22 January 2017, <https://www.cbc.ca/news/world/larry-chief-mouser-downing-street-1.3932276>

Sevin, Efe; ‘Twitter going viral: Twitter usage patterns in destination marketing and place branding’, 2013, 6, 3, 227-239

The Telegraph, ‘Meet the cats of Westminster: The Government mouser-in-chief’, 4 September 2017, <https://www.telegraph.co.uk/women/politics/meet-cats-westminster-government-mousers-in-chief/>

Wells, Peter, ‘UK Government cats’, 2016, <https://peterkwells.github.io/uk-government-cats/> [Last accessed: 27 February 2019]

Wells, Peter, ‘Creating a web of cats’, Lecture, *The Open Data Institute*, 25 November 2016, <https://theodi.org/event/friday-lunchtime-lecture-creating-a-web-of-cats/>

1. Porter, 2016 [↑](#footnote-ref-1)
2. Myrick, 2015, 171 [↑](#footnote-ref-2)
3. Podhovnick, 2016, 127 [↑](#footnote-ref-3)
4. Horton, 2016 [↑](#footnote-ref-4)
5. Davies, 1997 [↑](#footnote-ref-5)
6. The Telegraph, 2017 [↑](#footnote-ref-6)
7. Myrick, 2015, 174-5 [↑](#footnote-ref-7)
8. Scott, 2017 [↑](#footnote-ref-8)
9. Kogan, 2018 [↑](#footnote-ref-9)
10. Wells, 2016 [↑](#footnote-ref-10)
11. Open Data Institute, 2016 [↑](#footnote-ref-11)
12. Acer and Deguchi, 2013, 23 [↑](#footnote-ref-12)
13. Ibid, 24 [↑](#footnote-ref-13)
14. Sevin, 2013 [↑](#footnote-ref-14)
15. Du, Xie, Li, Zhu and Lim, 2013 [↑](#footnote-ref-15)
16. Komorowski, Huu and Deligiannis, 2017 [↑](#footnote-ref-16)
17. Ch’ng, 2015, 4: This definition was developed by D W McMillan and Chavis, 1986. [↑](#footnote-ref-17)