



Contents lists available at ScienceDirect

Government Information Quarterly

journal homepage: www.elsevier.com/locate/govinf

Linguistic and semantic factors in government e-petitions: A comparison between the United Kingdom and the United States of America

Stephen D. Clark^{a,*}, Nik Lomax^b

^a Leeds Institute for Data Analytics and School of Geography, University of Leeds, LEEDS LS2 9JT, United Kingdom

^b School of Geography and Leeds Institute for Data Analytics, University of Leeds, LEEDS LS2 9JT, United Kingdom



ABSTRACT

Many legislators around the world are offering the use of web based e-petitioning platforms to allow their electorate to influence government policy and action. A popular e-petition can gain much coverage, both in traditional media and social media. The task then becomes how to understand what features may make an e-petition popular and hence, potentially influential. One area of investigation is the linguistic and topical content of the supporting e-petition text. This study takes an existing methodology previously applied to the American government's e-petition platform and replicates the study for the United Kingdom's equivalent platform. This allows an insight into not only the United Kingdom's e-petition process but also a comparison with a similar platform. We find that when assessing an e-petition's popularity, the control variables are significant in both countries, e-petitions in the United Kingdom are more popular if some named entities are used in the text, and that topics are commonly more influential in America.

1. Introduction

Whilst the right to petition leaders and governments is an ancient right (Dodd, 2007; Fraser, 1961), the recent development of electronic petition (e-petition) platforms has raised the profile of petitioning in the social and political discourse (Leston-Bandeira, 2017). Many governments around the world now run their own e-petition platforms that allow citizens to highlight their concerns to legislators (Directorate-General for Internal Policies, 2015). This modern re-imagining of the petitioning process has the potential to illustrate the impact that information technology can have on the relationship between government and the electorate; with the potential for significant public influence on democratic practices, opening up another route of communication to the legislator that by-passes the "highest-level interest aggregator", e.g. the congressional representative or member of parliament (Taagepera, 1972).

Whilst the use of e-participation platforms have the potential to contribute to policy formulation and evaluation (Gil-Garcia, Pardo, & Luna-Reyes, 2018), the effectiveness of e-petitions in changing government policy is debatable (Bochel, 2012). For example in examining the United Kingdom (UK) e-petitioning platforms Hough (2012) and Wright (2015b) are generally sceptical, finding little evidence of e-petitions changing government policy, and the effectiveness in influencing gun-control laws in the United States of America (USA) is discussed in Dumas et al. (2015b) who find that e-petitions often promote divergent (pro- and anti-) gun control legislation options. In particular, Bochel

(2016) identify that there may be a gap between aspiration and reality for e-petitioners, with a need to manage the expectations of both e-petitioners and those who sign e-petitions. But it is undoubtedly the case that a popular e-petition can generate widespread public and media interest, helping to promote the agenda of its creator (Harrison et al., 2017).

Reasons for the popularity of e-petitions are not well understood. In some instances there can be an organised campaign to "get behind" a particular e-petition. Lee, Chen, and Huang (2013) identified this effect, with those having a strong political identification being disproportionately over-represented in the signatories to various e-petitions. Other campaigns can evolve more organically, with support for an e-petition growing over time via social media (Aragón et al., 2018; Margetts, John, Hale, & Yasseri, 2015).

A natural experiment study conducted by Hale, John, Margetts, and Yasseri (2018) attempted to see if the introduction of a list of recent trending e-petitions to the e-petition platform's home page impacted on the volume and the distribution of signatures. They found that the total number of signatures each day did not change much, but the trending e-petitions gained more signatures at the expense of those which did not feature on the list. However, it was still the case that the vast majority of e-petitions failed, with very few gaining anywhere near enough signatories to generate interest or prompt action (Yasseri, Hale, & Margetts, 2017).

A body of work has attempted to understand factors that influence the popularity of e-petitions through text analysis (Hagen et al., 2015).

* Corresponding author.

E-mail addresses: tra6sdc@leeds.ac.uk (S.D. Clark), N.M.Lomax@leeds.ac.uk (N. Lomax).

<https://doi.org/10.1016/j.giq.2020.101523>

Received 13 June 2019; Received in revised form 22 April 2020; Accepted 20 July 2020

0740-624X/ © 2020 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

These studies analyse the supporting text provided by the e-petitioner to infer how linguistic, semantic and topical factors are related to the popularity of the e-petition (Barats, Dister, Gambette, Leblanc, & Pérès, 2016). A major piece of analysis in this area was applied to the United States Federal Governments' e-petition platform "We the People" by Hagen et al. (2016). Their analysis is repeated in this article for the e-petitions hosted by the UK Parliament from May 2015 to September 2016. Their well cited study incorporates a wide range of possible influences on an e-petition's popularity, including control variables, linguistic characteristics, sentiment and the topic of the e-petition. Fortunately the format and intentions of the UK Parliament's e-petition platform are similar to the United States platform, so our analysis provides an insight into the operation of the UK Parliament's e-petition platform and will also illuminate and enable a reflection on the similarities to, and differences from, the original study.

Specifically, our analysis provides insight in to whether findings from one country or jurisdiction (in this case the USA) can be transferred to another, (here the UK) especially given similarities in language (Davies, 2005) (or dis-similarities, Algeo (1986)) and popular culture (Potts & Baker, 2012). Whilst it is difficult to form a priori hypothesis in the social sciences, where commonalities or differences do exist, there is the need to identify possible explanations for the findings. Possible reasons for differences may be institutional, influenced by the form of government; the esteem in which organisations or individuals are held; cultural norms particularly around social issues (e.g. religion, sexuality, guns); or linguistic, for example through the directness of language (Dunkerley & Robinson, 2016).

Commonly, little research in the social sciences is replicated in this manner, so the opportunity to compare and contrast findings, using an established and cited methodology and using similarly structured datasets is appealing. The code for our analysis has been made available, allowing researchers to further reproduce or build on our work, or adapt it for other jurisdictions.

2. Background

Since e-petitions exist on a web platform, the meta data concerning e-petitions has been widely used for secondary analysis in order to try and understand the e-petitioning process (Briassoulis, 2010; Contamin, Léonard, & Soubiran, 2017; Hagen, Harrison, & Dumas, 2018). In terms of the growth of individual e-petitions, an early study (in terms of the uptake of e-petition platforms) by Scott A Hale, Margetts, and Yasseri (2013) used daily web-scraped UK government e-petition data from February 2009 to March 2011 to examine how support for e-petitions grew. They found that an inflection point was reached after an e-petition reached the 500 signature threshold (during this period a government response was guaranteed when an e-petition reached this threshold). They also found that the distribution of signatures over time followed a recognised leptokurtic distribution and that the number of signatures gained on the first day was a significant factor in explaining the eventual number of signatures. This latter point was re-enforced in a follow-up study using more recent UK e-petition data that suggested an e-petition's fate in terms of popularity was decided in the first 24 h of its launch (Yasseri et al., 2017). Beyond the sheer popularity of some e-petitions, Puschmann, Bastos, and Schmidt (2016) investigate the behaviour of signatories of e-petitions, identifying classes of signers. These range from "Singletons" who signed just one e-petition through to the "Hyperactive" who contributed nearly 10% of signatures but made up just 0.1% of signatories. They were also able to examine this behaviour by different e-petition policy areas, with hyperactive signatories active foremost in the 'Labour' and 'Other' policy areas. Two companion articles by, Clark, Lomax, and Morris (2017) and Clark, Morris, and Lomax (2018) use e-petition data to, respectively, classify 'types' of Parliamentary constituency based on popular e-petition topics and to estimate the percentage of the European Union leave vote in each constituency. Both these studies exploit the richness of the

geographic detail at which signatory counts are made available to capture and typify the political sentiment within each Parliamentary constituency.

A new strand of analysis has been concerned with the influence that the language used in the e-petition text has on its popularity. The motivation behind our study is to replicate the methods and work flows of Hagen et al. (2016) and compare our findings to theirs. They report that "However to the best of our knowledge, no e-petition studies have addressed the impact of textual patterns on online campaigns." (p784) and instead refer to the more extensive literature to be found in the textual analysis of social media, in particular tweets. Here the methods used in the study by Hagen et al. (2016) are summarised before we introduce some subsequent studies which provide further insight in to the analysis of e-petition texts in other contexts.

Hagen et al. (2016) pose three research questions, supported by a review of studies that have used similar concepts in text analysis. The first research question (RQ1) asked "How will the linguistic variables of extremity, urgency, informativeness, request, internet activity, repetition and sentiment be related to petition signature accumulation and will they, as a block, account for a significant percentage of explained variance in petition signature accumulation?" (p 785). Many of the quantities in this block of measures were captured by using a look-up for particular words in the supporting text provided for each e-petition. For example, to measure Extremity a count was made of how often any of the words "much more", "extremely", "very", and "wonderful" appear in the text. Similar counts were conducted to measure: (1) Urgency, (2) Request, and (3) Internet Activity within the e-petition text. For Informativeness the total number of unique words were counted and for Repetition the number of words was divided by the total number of unique words. Sentiment was measured using the Stanford Sentiment Analyser (Socher et al., 2013), using a scale from 0 which denotes negative to 3 which denotes positive (and which accounts for modal shifters and intensifiers). All these variables were coded into binary variables depending on whether they were present at all (Extremity, Urgency, Request and Internet Activity); more frequent than the mean (Informativeness and Repetition) or fall in a range (strong negative or positive Sentiment coded as 1, neutral as 0). In their review of similar studies, they hypothesise that the more some of these features are present, the more popular an e-petition is likely to be (Intensity, Urgency, Informativeness, Sentiment and Internet Activity) or to just have a discernible effect that may be positive or negative (Repetition and Request). However, Hagen et al. (2016, p.785) also note that with all these features "... it is not always possible to specify the direction of that influence (positive or negative) ...".

The second question (RQ2) asked "Do semantic tagging variables (person, organization, and location), as a block, predict significant portions of explained variation in petition signature accumulation?" (p 786). To capture this block of influences the Stanford CoreNLP (Natural Language Processing) Named Entity Recognition (NER) tagger (Finkel, Grenager, & Manning, 2005) was used to count the number of persons, organisations or locations referenced by the e-petition. Their review of similar studies of the perception of named entities in text is non-comital on the direction of influence, primarily highlighting their potential to organise information and aide in decision making.

The final research question (RQ3) asked "Do naturally emergent topic variables, as a block, predict significant portions of explained variation in petition signature accumulation (25K and 100K petitions)?" (p 786). Capturing this block required the identification of data-driven topics that group together e-petitions. It was accomplished by using Latent Dirichlet Analysis (LDA) (Blei, Ng, & Jordan, 2000) to identify potential topics. LDA identifies topics by examining the co-occurrence of words in the texts and builds a probabilistic model to explain the distribution of words within topics and topics within texts. This establishes the nature of the topic by its association with certain words, and also to what degree each e-petition is associated with a topic (via a series of "affinity scores" which measure what proportion of the e-petition text is

likely to be attributable to each topic) (Hagen, 2018). The potential topics to include in a regression model were manually refined by only considering those that were coherent (on examination of the e-petitions within a topic, that most were on a similar theme) and also those that were relevant (in so far as they had a positive influence of the signature count prediction accuracy). As with NER, their review identified that topics have the potential to influence the perception of texts but that each topic's potential to influence e-petition popularity varies.

A fourth control block contained two measures of e-petition information: the number of signatories in the first 24 h of the e-petition; and the number of petitions opened on the same day as the e-petition. For their models they logged the non-dummy variables (except the affinity scores).

In their final model (Model 4 of Table 4) they found that the significant variables in predicting the popularity of an e-petition were: the number of signatures in the first 24 h; the number of petitions opened on the same day; Extremity; reference to a Person in the e-petition description; and the topics of Religion/Gay, Children, Secession, China, Awareness, Student visa, White genocide and Guns. The R_{adj}^2 was high at 32%.

Subsequently Porshnev (2018) applied part of the Hagen et al. (2016) method to the Russian Public Initiative e-petition platform using just the concepts of Informativeness, the presence of three terms ("to ban", "for all" and "Russia") and any revealed topics in the e-petitions. In the list of e-petitions he identified 20 potential topics using LDA. He also introduced a series of year dummies and a dummy for the geographic level at which the e-petition was relevant. Unusually with the Russian Public Initiative platform it is possible to sign both in favour and in opposition to each e-petition, so he was able to estimate two models, the Pro and the Against signatory counts. In the regression results, the year dummies were all significant and the significance of various topics varied between the Pro and Against models. Informativeness was not significant at the 10% level in the Pro model (which is the one most akin to those considered here) but was positive and significant for the Against model.

Research by Chen, Deng, Kwak, Elnoshokaty, and Wu (2019) also tried to use linguistic cues within Change.org e-petition texts to explain an e-petition's popularity. They formed hypothesis around the likely impact of the cognitive appeal (four hypotheses), the emotional appeal (two) and the moral appeal (two) of each e-petition on its popularity. Their regression equation consisted of these eight linguistic appeals, a pre-defined topic category chosen by the petitioner, plus a number of control variables including word count and effectiveness. They report that most of their eight linguistic appeal variables are significant at the 5% level (negative emotion wasn't significant) and all the topic variables were significant at the 10% level except for those in the Gay rights category.

Other studies use more complex machine learning algorithms to predict the popularity of e-petitions. Suh, Park, and Jeon (2010) use artificial neural networks and decision trees to forecast the daily likelihood that a South Korean e-petition will be a "nationwide matter". This is done by identifying keywords in each petition and using these keywords to place the e-petition into one of eight e-petition groups (topics) and then forecasting the trend in each group. They compare how well their models predict the speed at which each e-petition will become a national matter with estimates based on a manual assessment and judge that three of the groups achieve this status earlier than the manual assessment would suggest.

Focusing in on the distribution of e-petition topics, TeBlunthuis (2018) assigned individual change.org e-petitions to topics using LDA and measured the impact of petition density and specialization on popularity. They measured specialization according to the degree to which an e-petition references few topics or many, thus an e-petition with a large affinity score to one topic but low scores to others would be regarded as a specialist e-petition. They found that an inverted U shaped relationship existed between popularity and the density of e-

petitions within a topic, that topics with a moderate number of e-petitions performed well, but topics with few or many e-petitions did less well. They also found that more specialist or niche e-petitions did not outperform those that were more generalist. There may therefore be a competition for signatories in an e-petition "market", with a generalist e-petition in the company of a reasonable number of similar e-petitions better at attracting signatures.

It is clear from the initial assertion from Hagen et al. (2016) and a search for subsequent literature that the area of linguistics and semantic analysis of e-petition texts is a new, under researched field, but it does have the potential to provide insights into this relatively new form of political engagement.

3. Materials

The current version of the UK Parliament's e-petition platform came into operation following the May 2015 General Election (Houses of Parliament, 2017). Previously the platform was hosted within the office that reported to the Prime Minister (Wright, 2015a). British citizens and UK residents can create an e-petition and to get the e-petition started requires just five people to support it. After the e-petition is checked to ensure that it meets the standards for e-petitions (in-particular that it does not replicate an existing e-petition), it is published on the UK Parliament's e-petition platform (see Supplementary Fig. S1 for a screen shot of an e-petition's page). British citizens (in the UK or overseas) and non-British UK residents can then sign the e-petition. To sign an e-petition the user is required to confirm their citizenship or residence, supply their name and an email address. An email is then sent to this address and the signing takes place when the user clicks on a link provided in the email. There is no facility or requirement to be a registered user. At 10,000 signatures the e-petition gets a response from the government and at 100,000 signatures the e-petition will be considered for a debate in Parliament. All e-petitions stay open for 6 months, but can be closed before this time if the current Government steps down. This is the case for the e-petitions used in this study, where an early General Election in June 2017 closed all the open e-petitions at that time.

Since individuals do not need to register to use the platform and the signatories' identity or email addresses are not made public it is not possible to track individuals across e-petitions. Also there is no facility for the signatory to leave comments on the e-petition's page (this is possible with some platforms, e.g. lapetition.be (Contamin et al., 2017)).

A number of e-petitions have become talking points in both the mainstream and social media. Of the e-petitions included in this study, an e-petition to define the rules for a second EU referendum and another to prevent a state visit for President Donald Trump gained over 4.1 million and 1.8 million signatures respectively. Both of these e-petitions received extensive coverage in the media, prompting comment from senior politicians (Slawson, 2016, 2017). However, most e-petitions are less popular. Of the 7828 e-petition that were open for at least 180 days, the median number of signatures is just 49 (in this positively skewed distribution, the mean is much higher at 3160 signatures).

3.1. Data acquisition

In this study the approach of Hagen et al. (2016) will be replicated for the UK Parliament's e-petition archive from May 2015 to September 2016. Most of the data required is taken from the archived version of the UK Parliaments' e-petition web site that makes e-petition data available as a JSON file. An R (R Core Team, 2017) script¹ is used to

¹ The R script and the .RData workspace are available as Supplementary material.

download the data associated with e-petitions created between May 2015 to June 2017, providing data on 10,949 of the 10,950 available e-petitions (the data for one petition is corrupted). Unfortunately the number of signatures within the first 24 h is not provided in these JSON files. Instead this information has been sourced via the Oxford Internet Institute who maintained a web crawler to archive the number of signatures for each e-petition every hour (Hale, Margetts, & Yasseri, 2019). This information is however only available until 12 September 2016, hence only the e-petitions opened in the first 18 months of the May 2015 to June 2017 Parliament are used. The truncation of these data is justified since many studies have highlighted the importance of this first 24 h period on an e-petition's eventual popularity. Of the 10,949 e-petitions, those started in the first 18 months and for which we have a count of signatures in the first 24 h, provides 7205 e-petitions, removing around a third. A further 23 e-petitions were removed because the only available time for the initial number of signatures was not between 24 ± 3 h after it was opened. One e-petition had a number of signatures in the first 24 h much greater than its eventual total after 6 months and was therefore removed. These exclusions leave 7181 e-petitions for analysis, of which 3590 are randomly selected for the training of the topic model and the remaining 3591 used for estimating the regression model. The topicmodels package (Grun & Hornik, 2011) is used to conduct the LDA and the coreNLP package (Manning et al., 2014) is used to measure the sentiment and count the number of named entities referenced in the e-petition.

3.2. Topic modelling

The corpus for each e-petition is built from the action text (usually a single sentence or line of text) and the more copious additional text and background text information. The corpus is treated to remove spurious white space, punctuation and stopwords (also the word "government" is removed since it is used often with little specificity). The characters are converted to lower case and the words are then stemmed to a common root. At this point the number of words and the number of unique words are recorded to provide the Informativeness and Repetition measures. This full corpus consists of over 28,260 words. At this stage the corpus is randomly split into two equal parts. One part is a training corpus used solely to identify potential topics within the e-petitions. This part is not used for regression modelling so does not need any information in regards to the e-petitions' linguistics characteristics, sentiment or named entity recognition and the number of signatories is irrelevant. The second testing part is used for regression modelling and requires all these pieces of information and also their topic, using the topic model developed using the training corpus.

To decide the number of topics, a range of candidate topic numbers are fitted using the Gibbs method for LDA with the training data, using a burn-in of 2000 iterations, followed by 2000 further iterations, keeping every 50th iteration and the evaluation is carried out using a 10-fold cross validation. The scree plots for perplexity and for the log likelihood are shown in Fig. 1.

These plots suggest that there are between 25 and 30 topics within the e-petitions. For the 25 topic solution, the beta values for words with the eight highest betas in each of these topics are shown in Fig. 2. An examination of these words and the subject matter of the top twenty e-petitions within each topic (by affinity score) (see Supplementary Table S1) provides coherent topic descriptions for 21 of the topics (Hagen et al. (2016) also had 25 candidate topics, 18 of which are coherent). At this point it is worth mentioning that there is no requirement here for a value judgement as to whether the e-petition is pro- or anti- the assigned topic. The trained LDA 25 topic model is then used to predict the most likely topic for each e-petition in the testing data. The top twenty e-petitions within each topic for these testing data is provided in Supplementary Table S2 and it is clear that the trained topic model has performed well in identifying the most likely topic for this unseen testing data set.

3.3. Descriptive statistics for variables

To see how the nature of our data compares with that of Hagen et al., 2016 a version of their Table 3 is produced here using our data, see our Table 1. This includes their means for comparison.

For this study there are over twice as many e-petitions in the test set compared to Hagen et al. (2016). The mean number of signatures and the number of signatures in the first 24 h is less in the UK (however the residential population of the UK is also less than the USA). The mean number of e-petitions opened each day is however greater in the UK, but this may be a result of differences in how the two platforms operate. In the UK a petitioner is required to supply the email addresses of just 5 supporters and once an e-petition has been checked to ensure it meets the various standards it goes live. In the USA the petitioner is required to gather 150 signatures before the e-petition goes live on the platform, a much higher threshold. The linguistic style variables compare well, except with Request; in the UK the action text for e-petition tends not to exhort others to share or spread the e-petition, so the value here is lower. UK e-petitions are also less likely than their USA equivalents to contain references to named entities. In regards to topics, the sum of affinity scores for each topic is similar at 4.0%. The maximum affinities for each topic in the UK is lower than those in the USA, which suggests that the UK e-petitions are less strongly "topiced" than those in the USA. When each e-petition is assigned to the one topic with the highest affinity score (the column labelled n), the Referendum, Education and Medical Treatments topic are very common, whilst the incoherent topics are, generally, least common.

The distribution of the logged number of signatures received for each topic by the highest affinity score, is shown in Fig. 3, ordered by the median, (here both the training and test data sets are combined). E-petitions that call on the government to support various medical treatments are popular, as are those around asking government to influence the actions of local government and promote animal welfare. Surprisingly the e-petitions concerned with the process and outcomes of Referendums (both to leave the EU and for Scottish independence) are the least popular, however there are many of these e-petitions available for signing (with the competition potentially diluting the impact of many) and one of these e-petitions did gain by far the most signatures.

3.4. Topic relevance

Hagen et al., 2016 further sub-setted the coherent topics to remove those that were not relevant, as described in their Appendix 1. This task ensures that only topics that are relevant (and not just coherent) are included in the final model. A topic is defined as relevant if its removal from a model of the number of signatures against all coherent topics does not decrease the goodness-of-fit of the model (i.e. retaining the topic makes a better model).² This is achieved using a 10-fold cross classification technique with the training data, and dropping as irrelevant those topics that did not decrease the mean square error relative to a full model. Applying this further selection in our study removed seven topics, leaving just 14 coherent and relevant topics from the original 25 (Hagen et al. (2016) finished with 15 of the original 25). In their Appendix 2 they also compare the topic parameter estimated from the trained and the testing data. The purpose of this analysis is to ensure results derived from one sample of data are generalisable to a different set of data. To see if this is the case, all 14 coherent and relevant topics are included in a regression model estimated separately using the training and testing data. If the results are generalisable then the parameter estimates from these two models should be, in the context of their estimated standard errors, similar. This degree of similarity is shown here in Fig. S2 of the supplementary material. In all cases the parameter confidence intervals based on the training and the test data overlapped for each topic, meaning that the predictive performance associated with various topics is not likely to be influenced by the original training/testing split of e-petitions.

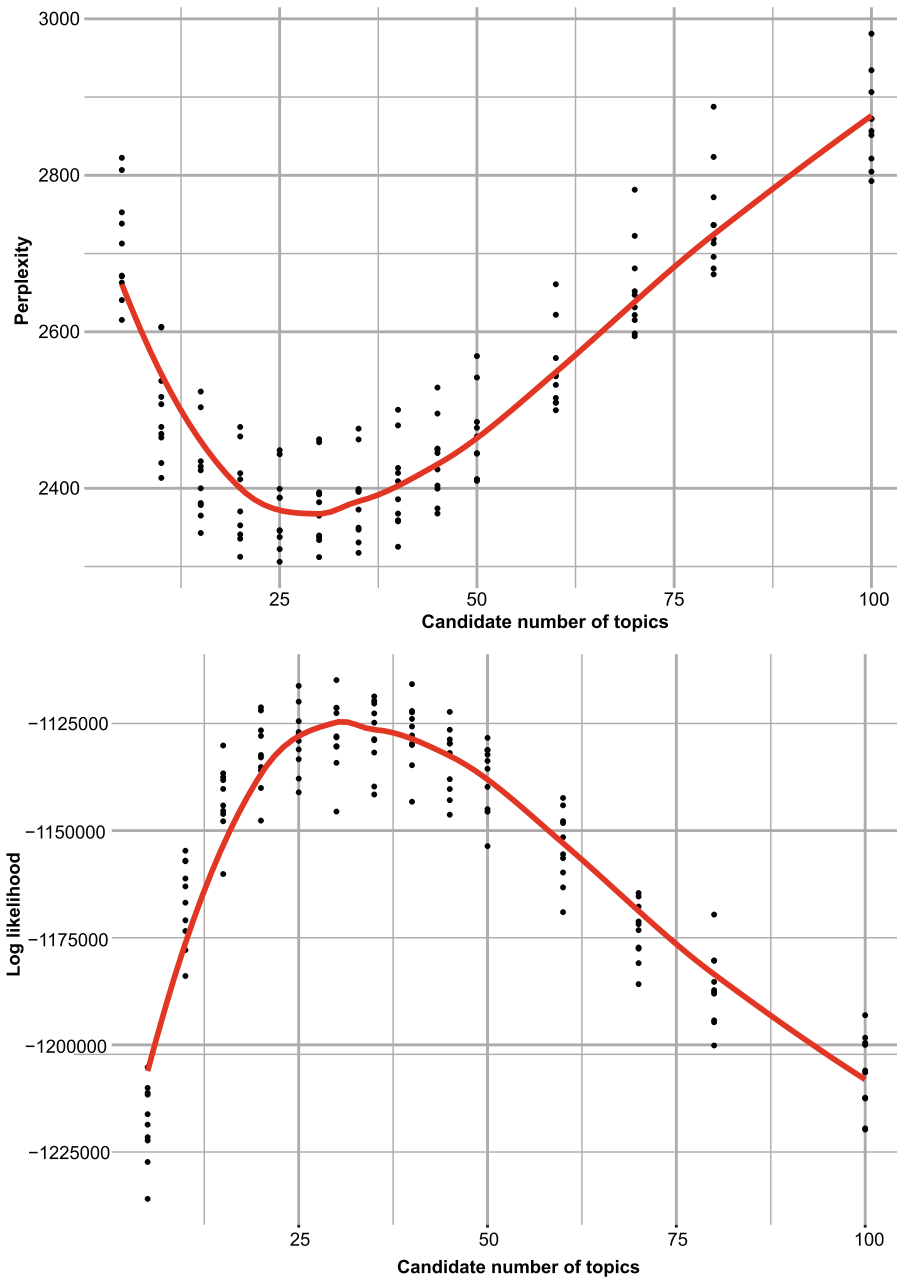


Fig. 1. Scree plots of perplexity (top) and log likelihood (bottom) to estimate the number of topics.

4. Results

The results from the regression equations are shown in Table 2. The Control model includes just the control terms; Linguistic introduces the linguistic and semantic terms; and NER introduces the three named entities into the model. The Full model reports the full regression results, including all topic affinities. For comparison the results from Hagen et al., 2016, Table 4, Model 4, are provided in the final column and their topics in the footnote. The goodness of fit for the models is reported via the R_{adj}^2 values, and the goodness of fit for the full regression model estimated on UK data is higher, at 0.55, than that reported for USA models, at 0.33. For the UK the increase in R_{adj}^2 as blocks of variables are added is small, but assessed using an analysis of variance, all these increases are significant at the 0.1% level.

The number of signatures received in the first 24 h is significant for all the regression models, with its magnitude being reduced only slightly as further terms are introduced. The value is much larger than

that reported in Hagen et al., 2016. In the UK the coefficient for the number e-petitions opened on the same day is negative and significant, whilst the equivalent co-efficient is positive and significant for the USA.

For the linguistic factors, only e-petitions that contain material related to internet activities significantly increases the number of signatures. In the USA these coefficients, except for Extremity, are also not significant. In the UK, some of the coefficients that count references to named entities have a significant and positive effect, whilst in the USA only the person co-efficient is significant and it is negative. In the USA, 8 of the 15 topic variables are significant whilst in the UK there are only 3 topics with a significant impact on signatures: Medical Treatments, Animal Welfare and (the treatment of the) Vulnerable. The Animal Welfare topic is particularly positive and significant. This relatively low number of significant topics is also reflected in the small increase in R_{adj}^2 (of 0.01) when topics are included (Hagen et al. (2016) reported an 0.08 increase).

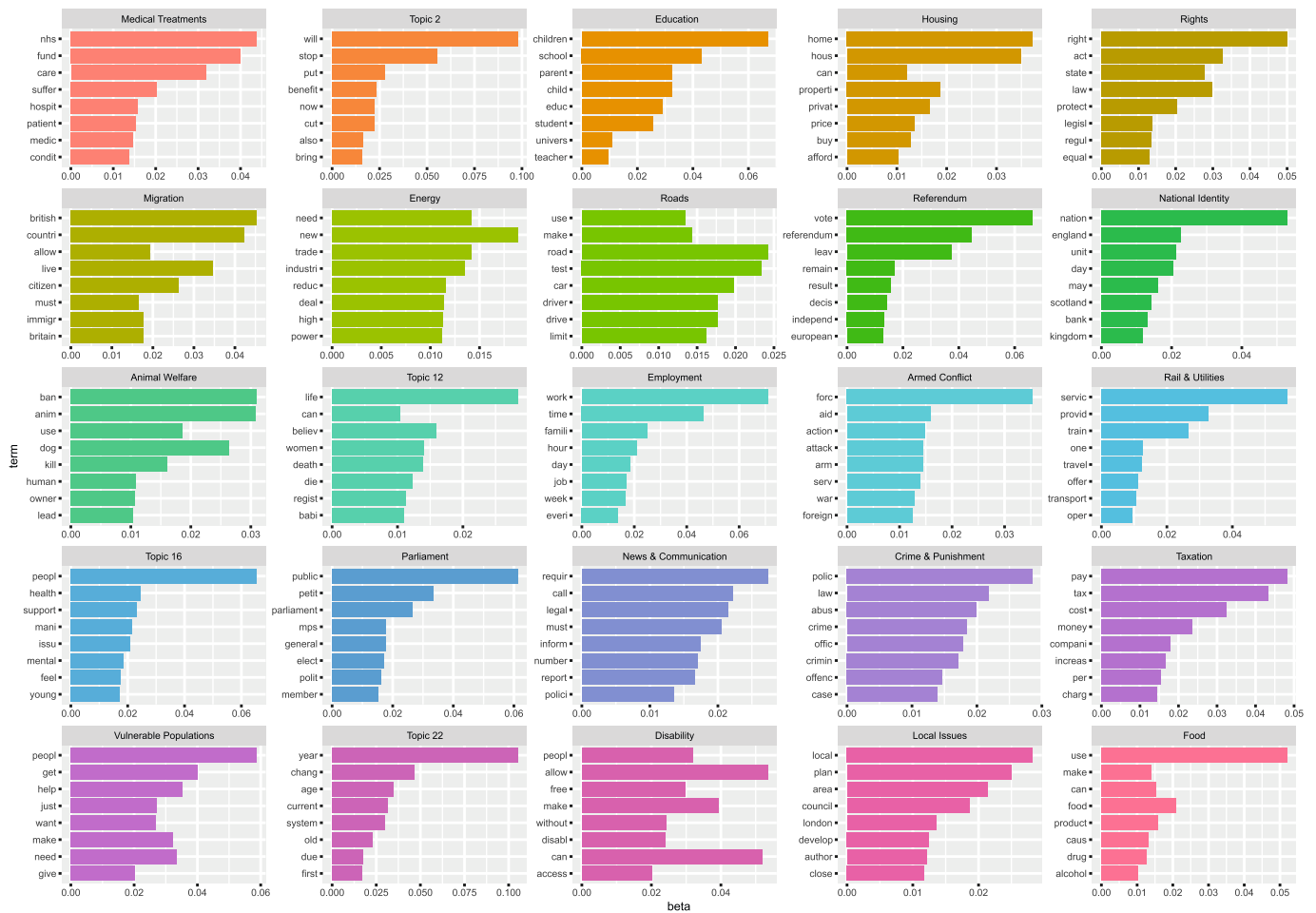


Fig. 2. Words with the top ten beta values in each topic and topic titles.

5. Discussion

In this study, the methodology outlined in Hagen et al. (2016) has been applied to an equivalent data set constructed from the UK Parliament's e-petition platform. To make the comparison as close as possible the same linguistic cues, sentiments and named recognitions software are used.

5.1. Control variables

In terms of the regression results, both studies have identified that the number of signatures received in the first 24 h is a significant indicator of its eventual level of popularity. However, since most e-petitions “fail”, with over a half of the e-petitions attracting less than 50 signatures over a 6 month period, an initial low count of signatures is an obvious candidate for a significant variable. This feature is apparent when the number of signatures received in the first 24 h is tabulated against the eventual number of signatures (see Table 3). Of the just under 3700 e-petitions that gain 15 or fewer signatures in the first 24 h, less than 1% reach the threshold for either a government response (10,000) or consideration for a debate in Parliament (100,000). However, a low initial number of signatures is not a death knell for an e-petition, nine e-petitions that did poorly initially did reach this challenging upper threshold of more than 100,000 signatures by the end of the six-month period.

The more e-petitions that are opened on the same day, the lower the number of signatures. This suggests there is evidence of a competition for signatures in the UK (recall that on average more e-petitions are opened each day in the UK than the USA, but note that the USA requires

a higher threshold before an e-petition is listed). Governmental authority is more centralised in the UK than the USA (Booth 2015), meaning that the UK Parliament's e-petition platform is the only natural, government sanctioned, mechanism to raise a concern. This creates a diversity and volume of e-petitions that can “crowd out the market”, with many e-petitions competing for public interest on a wide range of subjects. Also, the UK e-petition platform lists on its front page the top three trending e-petitions during the last hour and an e-petition placed in this short list is likely to generate more signatures, particularly in the crucial first 24 h of the e-petition. The greater the number of new e-petitions, the less likely that one of these limited number of slots will be available. This emphasis on initial popularity is seen in other markets that chart popularity, such as with books (Sorensen, 2004) and music albums (Asai, 2009). In the USA the impact of having many e-petitions opened on the same day is however positive, which suggests less of a competition effect between e-petitions, with less viable USA e-petitions not appearing on the platform by virtue of the 150 signature threshold. On the “We the People” platform, the front page also lists all open e-petitions in decreasing order of overall popularity, not just popularity in the last hour. Thus a popular e-petition will re-enforce its popularity over a longer period by heading this list.

5.2. Linguistic and semantic variables (research question RQ1)

The summary statistics for the linguistic and semantic variables (Table 1) shows a similarity in their values between the UK and the USA, with the exception of the Request variable. The UK e-petitions tend to make fewer requests for people to sign the e-petition in the headline action text than they do in the USA. This may be a reflection of

Table 1
Descriptive statistics of all variables in the test set (3,591 here and 1671 for Hagen et al., 2016).

	Table 3, mean from Hagen et al., (2016)	Mean	Std. dev.	n	Min	Max
Signatures count ^a	7.59	4.51	2.16	3591	1.79	15.24
Petitions created ^a	1.69	3.46	0.66	3591	0	4.81
First 24 h ^a	4.62	3.12	1.40	3591	0	11.39
Informativeness	0.69	0.51	0.50	3591	0	1
Internet activity	0.10	0.13	0.33	3591	0	1
Request	0.19	0.02	0.14	3591	0	1
Extremity	0.19	0.18	0.39	3591	0	1
Repetition	0.44	0.44	0.50	3591	0	1
Urgency	0.56	0.55	0.50	3591	0	1
Sentiment	0.21	0.16	0.37	3591	0	1
Person ^a	0.50	0.17	0.42	3591	0	2.944
Location ^a	0.72	0.50	0.59	3591	0	2.708
Organisation ^a	0.67	0.49	0.66	3591	0	2.996
Animal welfare		0.038	0.033	157	0.012	0.366
Armed conflict		0.038	0.031	179	0.013	0.361
Crime & punishment		0.039	0.031	168	0.013	0.287
Disability		0.038	0.021	50	0.013	0.261
Education		0.042	0.036	279	0.013	0.296
Employment		0.040	0.029	117	0.013	0.406
Energy		0.038	0.028	121	0.013	0.299
Food		0.039	0.032	154	0.013	0.317
Housing		0.037	0.026	105	0.013	0.325
Local issues		0.039	0.027	116	0.013	0.297
Medical treatments		0.042	0.036	251	0.013	0.316
Migration		0.040	0.030	183	0.014	0.278
National identity		0.037	0.024	85	0.014	0.316
News & communication		0.041	0.024	69	0.014	0.274
Parliament		0.044	0.035	216	0.014	0.308
Rail & utilities		0.038	0.026	98	0.013	0.317
Referendum		0.046	0.043	321	0.012	0.359
Rights		0.040	0.026	114	0.013	0.294
Roads		0.042	0.042	227	0.013	0.362
Taxation		0.045	0.036	212	0.014	0.343
Vulnerable		0.043	0.025	81	0.014	0.209
Topic 2		0.039	0.019	53	0.014	0.154
Topic 12		0.037	0.023	81	0.013	0.264
Topic 16		0.040	0.027	108	0.013	0.306
Topic 22		0.039	0.022	46	0.012	0.187

^a These variables are on the log scale.

the less ‘direct’ nature of British society, with an underlying, unstated, assumption that the e-petitioner wants people to sign. American English in contrast can be more direct and transparent (Dunkerley & Robinson, 2016; Grainger & Mills, 2016), with an e-petitioner being more comfortable with an upfront request for the reader to sign the e-petition.

In the UK, providing additional material to be accessed via the internet is the only significant variable and this increases the number of signatures. E-petitions mentioning the internet, or providing links, could benefit from promotion by online communities (Sheppard, 2015) or via social media (Dumas et al., 2015a), creating a positive “bandwagon” effect. Also on a practical basis, since signing takes place online it may be the case that people either follow through to the internet resource linked and agree with its premise or its presence in the text lends some authenticity or added weight to the e-petition which inclines the individual to believe the e-petition is worth signing. The USA also has a positive estimate for this term, but it is not significant. In the USA only the degree of Extremity in the e-petition text significantly reduces the number of signatures, whilst in the UK Extremity increases this number. Of the remaining terms that are insignificant, Urgency and Request, both have the same sign for both the UK and USA, whilst Informativeness and Repetition (both word count terms) have different signs. Sentiment is not significant in both countries. For the UK, the mean sentiment across all the test e-petitions is measured on the scale as 1.441, whilst the standard deviation is 0.306. Few e-petitions (255)

have a sentiment value at or above 2.0 and slightly more are at or below 1.0 (356), meaning that 83% of e-petitions are considered neutral. This high percentage of e-petition which are considered neutral is perhaps somewhat of a surprise. The purpose of an e-petition is to excite a response from the reader, and a neutral tone to the text is unlikely to achieve this outcome (Berger & Milkman, 2012). Many e-petitions contain a mixture of both negative statements (e.g. attacking a government decision) and positive sentiments (e.g. suggesting a more palatable alternative course of action) but these more “extreme” sentiments are averaged out to produce a more neutral tone for the entire text. This lack of variation in the sentiment scale may explain its poor explanatory power.

In the model reported by Porshnev (2018), he found that in his model for Pro signatures in support of e-petitions, the linguistics measure of Informativeness was not significant at the 5% level and of the 15 topic variables, seven were significant at the 5% level and five were significant at the 1% level. The R_{adj}^2 value was low at just 0.15.

5.3. Named entities (research question RQ2)

Named entities are invoked less often in UK e-petitions compared to those in the USA, thus named entities are rarer in UK e-petition text. Given this relative rarity, referencing an actual entity, be it a person or an organisation significantly increases the number of signatures in the UK. The presence of such entities provides informational cues to the reader, particularly the ‘aimless petitioners’ who “... will be more easily shaped by information cues” (Hale et al., 2018) page 16). The presence of the named entities in the e-petition text may also “humanises” the e-petition, so that individuals react to emotional cues within the text and this transforms the intent of the e-petition from an abstract concept into something particular and specific (“The Appeal of the Narrow”, Hersh and Schaffner (2018)). Additionally the popularity of these e-petitions that reference persons or organisations may point to a successful campaign to promote the issue behind the e-petition.

Referencing a location will also increase the number of signatures, but not significantly so, which may point to these e-petition being too niche, concerned with a local issue and therefore having a smaller pool of potential signatories to call upon. For example, Clark et al. (2017) find that specific, localised e-petitions (e.g. signatories for a petition to save the steel industry are focused primarily in just two constituencies) are less useful in their classification algorithm because they are not representative of the wider electorate. In the USA however, there is only a negative effect associated with the naming of people. One explanation for this negative effect in the USA is that such terms might lead to fairly specific e-petitions that would struggle to gain more general popular support. In the UK, a different mechanism is clearly present in regards to the naming of entities in e-petition text than in the USA.

5.4. Topics (research question RQ3)

Whilst the number of topics discovered is similar in the two sets of e-petitions, the UK e-petitions tend to less strongly topiced than those in the USA, with lower maximum affinities (see Table 1). This weakness for topics means that many fewer topics have a significant influence on the number of signatures in the UK than in the USA. The largest influence is with e-petitions around Animal Welfare, and the impact is large. It is often said that “England is a nation of animal lovers” (Egan, 2014) (page 71) and this may be a manifestation of this affection for animals and their welfare. The only other e-petition topic that attracts a significantly higher number of signatures are those in support of making various medical treatments available, either through the National Health Service or by legalising, for therapeutic use, various currently banned substances. Generally in society, support for the National Health Service remains strong in the UK and its continued funding to support universal care, irrespective of means or location, is

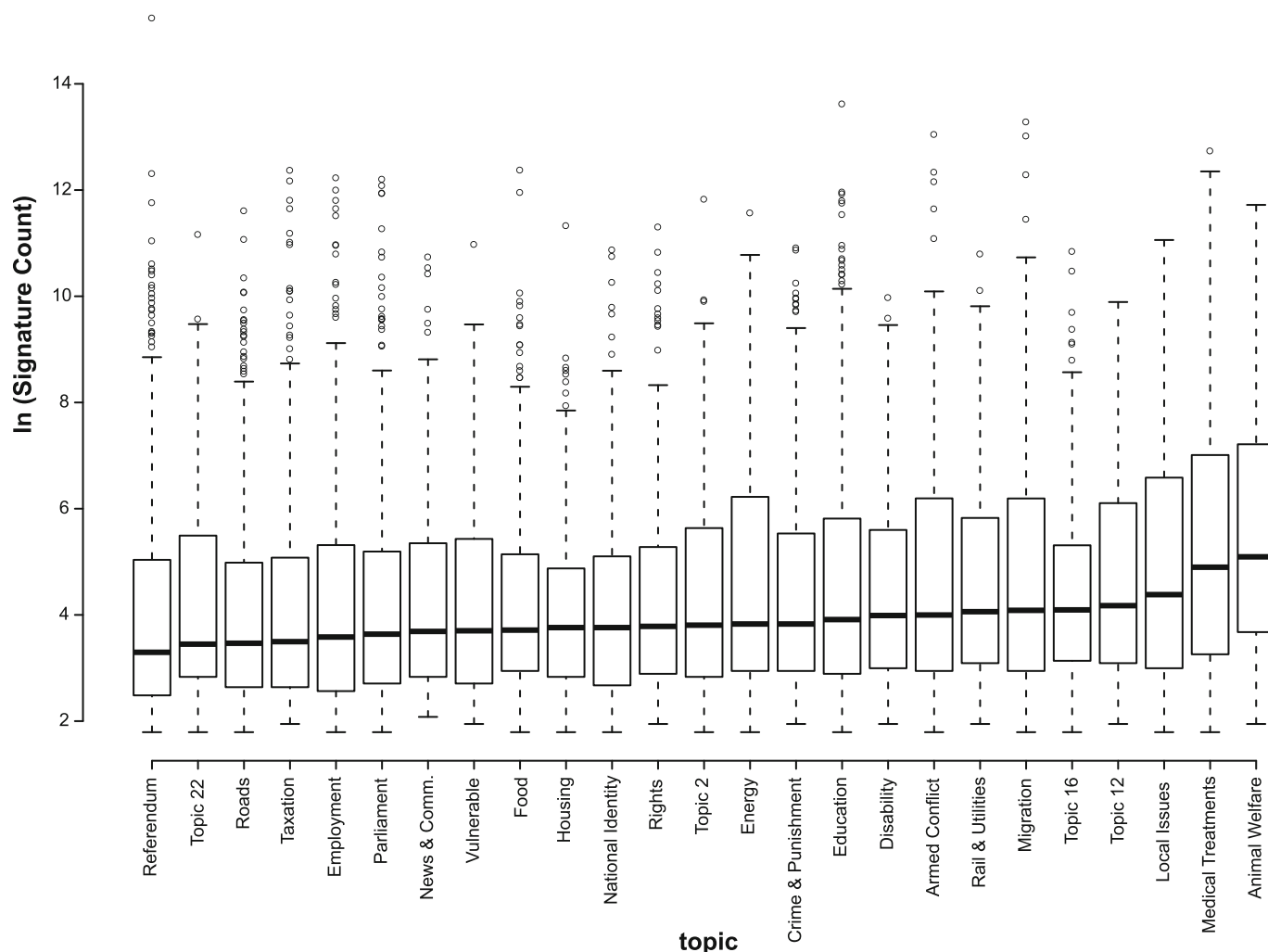


Fig. 3. Distribution of the logged number of signatures by topic group ($n = 10,949$).
 Note: * These variables are on the log scale.

high in people's priorities (Cream, Maguire, & Robertson, 2018; Wellings, 2018).

A surprising result is that whilst perhaps the most contentious political topic of Referendums did have an estimated positive effect on the number of signatures, this was not significant. These e-petitions referenced either the aftermath of the Scottish Independence referendum of September 2014 (Mullen, 2014) or the UK's referendum to leave the EU in June 2016 (Jackson, Thorsen, & Wring, 2016) and were the most numerous of topics to be found in the testing data set. It is plausible that this plethora of potential e-petitions on this topic dilutes the popularity of most of them – a competition exists for signatures within the topic, lowering the number of signatures for any one e-petition. Relative to the incoherent or irrelevant topics, the only other positive effects are associated with the topics of Employment, Taxation and Food.

The level of concern for animals reported above does not appear to apply to the Vulnerable in society such as the homeless and destitute, as e-petitions concerned with the welfare of these groups attract significantly fewer signatures (O'Neil, Pineau, Kendall-Taylor, Volmert, & Stevens, 2017). E-petitions around the topic of Disability are also less popular.

While it is difficult to make direct comparisons between the influence of the topics reported in Hagen et al. (2016) and those available in the UK e-petition data, some topics have the potential to overlap. In the USA, the coherent and relevant topics of cancer screening and marijuana correspond with the appeal in some of the UK Medical

Treatments e-petitions to enhance the screening provision for certain conditions or to introduce medical marijuana to help with the symptoms of some illnesses. The vexing issues in the UK about Scottish independence and the relationship with continental Europe expressed in the various Referendum topped e-petitions also aligns well with the Secession topic in the USA. Finally, the USA topics referencing the care of veterans and the status of the military also are present in UK e-petitions that are concerned with armed-conflicts.

5.5. UK findings

Aside from the comparative aspect of this study, for the UK 21 coherent topics are discovered within the e-petitions opened during the early months of the May 2015 to June 2017 UK Parliament. These topics are insightful in their own right (Anthony & Haworth, 2020). Whilst some e-petitioning systems require or allow a petitioner to select a category for their e-petition, the current UK platform does not (previous incarnations of the UK government e-petition platform did require the petitioner to select one of 16 categories (Hale et al., 2013) or nominate a responsible government department (Yasseri et al., 2017)). Given the volume of e-petitions submitted and eventually hosted by the platform it is difficult to manually monitor trends around submitted and approved e-petitions that might indicate popular topics. With the LDA developed here, it appears feasible to automatically and consistently categorise e-petitions as they are submitted or hosted. If this were to be

Table 2
Results of regressions.

	Control	Linguistic	NER	Full	Table 4, Model 4, Hagen et al. (2016)
(Intercept)	1.80 *** (0.15)	1.65 *** (0.15)	1.59 *** (0.15)	1.59 *** (0.25)	6.30 *** (0.42)
Ln first 24 hours	1.12 *** (0.02)	1.11 *** (0.02)	1.10 *** (0.02)	1.09 *** (0.02)	0.31 *** (0.01)
Ln frequency	-0.22 *** (0.04)	-0.21 *** (0.04)	-0.22 *** (0.04)	-0.21 *** (0.04)	0.20 *** (0.03)
Informativeness		0.14 ** (0.05)	0.08 (0.05)	0.08 (0.05)	-0.009 (0.07)
Internet		0.21 ** (0.07)	0.17 ** (0.07)	0.15 ** (0.07)	0.01 (0.10)
Extremity		0.04 (0.06)	0.06 (0.06)	0.06 (0.06)	-0.17 ** (0.08)
Repetition		0.03 (0.05)	0.05 (0.05)	0.05 (0.05)	-0.06 (0.06)
Urgency		0.06 (0.05)	0.06 (0.05)	0.06 (0.05)	0.08 (0.06)
Request		-0.27 (0.17)	-0.25 (0.17)	-0.19 (0.17)	-0.06 (0.08)
Sentiment		-0.05 (0.07)	-0.04 (0.07)	-0.03 (0.07)	-0.02 (0.07)
Ln person			0.12 ** (0.06)	0.13 ** (0.06)	-0.11 ** (0.05)
Ln location			0.09 ** (0.04)	0.08 * (0.05)	0.05 (0.05)
Ln organisation			0.12 ** (0.04)	0.12 ** (0.04)	0.02 (0.05)
Medical treatments				1.55 ** (0.72)	
Housing				-0.23 (0.97)	
Roads				-0.98 (0.63)	
Referendum				0.55 (0.67)	
National identity				-0.15 (1.07)	
Animal welfare				4.22 *** (0.78)	
Training				0.10 (0.89)	
Armed conflict				-0.13 (0.87)	
Parliament				-0.77 (0.77)	
Taxation				0.39 (0.72)	
Vulnerable populations				-2.53 ** (1.02)	
Disability				-1.32 (1.18)	
Local issues				-0.66 (0.96)	
Food				0.25 (0.81)	
N	3591	3591	3591	3591	1671
R _{adj} ²	0.54	0.54	0.54	0.55	0.33
R _{adj} ² change		0.002***	0.003***	0.005***	

Hagen et al. (2016) identified the topics of: Veteran; Religion_{gay}***; Children***; Investigation; Marijuana; Sentence; Cancer_{research}; Secession***; China***; Awareness; Student_{visa}***; Military; National Park; White_{genocide}***; and Gun**.

- *** p < 0.001.
- ** p < 0.05.
- * p < 0.1.

done with the e-petitions that are currently active on the UK Parliament's platform, and with those that will be hosted by over time, periodic re-training may be able to identify emerging topics or discard topics that are no longer relevant (much as Vidgen and Yasseri (2020) examined the temporal dynamics in these same data).

Table 3
Cross tabulation of e-petition popularity during first 24 h vs eventual popularity.

First 24 h	Less than 21	21 to 50	50 to 300	300 to 10,000	10,000 to 100,000	More than 100,000
Less than 9	1388	313	213	138	15	2
9 to 15	661	497	291	137	25	7
16 to 40	83	724	659	283	34	5
More than 40	0	4	560	897	215	30

Since these e-petitions are made to governmental authorities to achieve some goal, the nature of this authority may have an influence. The UK still has an organic form of government that follows a model with a centre-periphery structure (in spite of the recent devolution of some powers to Scotland, Wales and Northern Ireland) whilst the USA political system is designed to be more federal (Elazar, 2016). Thus e-petitions on local niche matters are pertinent for the UK platform but less so for the USA platform, where alternative layers of government responsibility are available to the petitioner. So in the UK context an e-petition that sought to raise a local issue concerning an individual or organisation might attract few signatures overall, thereby diluting the contribution of similarly topiced e-petitions, but if there were a concentration of these signatures in a particular cluster of Parliamentary constituencies this may cause it to be considered a local success (for examples of these, see Fig. 1 of, Clark et al. (2017)). This spatial concentration of support can be readily identified, since on the UK Government's platform, counts of signatures by constituency are published. Additionally, dedicated e-petitioning platforms for local government authorities that would provide a more neighbourly avenue of redress and publicity for e-petitioners are an active area of consideration (Bochel & Bochel, 2016).

5.6. Limitations

In this study we have chosen to follow the methods used by Hagen et al. (2016) to facilitate the comparison. In doing so, we accept the potential limitations as expressed in that study. These were that the NER package was trained using news articles rather than e-petition texts and that no attempt was made to validate the model against social events, e.g. a mass-shooting or the activities of certain religious groups. Whilst other sentiment analysers are available (Gonçalves, Araújo, Benevenuto, & Cha, 2013; Jongeling, Datta, & Serebrenik, 2015) we retained the use of the Stanford Sentiment Analyser to maintain alignment with the tools used by Hagen et al. (2016).

There is scope to further develop these methods in the future, for example, the CoreNLP package in R provides information on the parts of speech (verbs, nouns, adverb etc.) present in the text, and these, framed around sufficient hypothesis could be used to provide additional lexicographical insight as to what influences, if any, they might have on an e-petition's popularity. Hagen et al. (2016) also chose to codify sentiment as either neutral or not, with no differentiation between an e-petition that expresses largely positive sentiments verses one that expresses negative sentiment. There is an argument that one direction of sentiment may be more energising than the other, for example in the context of social media, Salathé, Vu, Khandelwal, and Hunter (2013) found that with messages around vaccinations, negative sentiments were "contagious" whilst positive sentiment not so. Here, Section 5.2 highlighted that an e-petition that contains a mix of both positive and negative sentiments could average out to suggest a more neutral sentiment, which loses some of the richness of the text. To capture the variation in sentiment, a statistical measure of the variability in sentiment throughout the text could be tested.

The analysis of a corpus as large as the one considered here presents some challenges. The texts for such e-petitions, having been created by

a wide and diverse range of authors, are unstructured which can be difficult for algorithms to interpret. However, the automated techniques available and used here are well established and appear to be able to extract a number of coherent topics.

6. Conclusions

This study has demonstrated that it is possible to replicate the methodology used in a study of the linguistic and topicality relevance of e-petition text derived for one country to another. Whilst the study by Porshnev (2018) using Russian government e-petition attempted this in part, ours is the first replication that is largely faithful to that of Hagen et al. (2016). This work has been facilitated by the willingness of those hosting such platforms to make their data freely available for research and study.

A comparison of the modelling results from the UK and the USA reveals little in the way of commonality. The strongest agreement is with the Control variable that counts the number of signatures in the first 24 h, with this term having a positive and significant impact, more so for the UK. The other Control variable, the number of e-petition opened on the same day, is significant in both countries but with opposite signs. In the UK there appears to a competition for signatures, and the consequential right to claim a valuable trending spot on the front page of the e-petition platform. Looking at linguistic terms, very few such terms are significant for either the UK or the USA, references to the internet increase signatures in the UK whilst Extremity of content decreases popularity in the USA. The presence of named entities always increases the number of signatures in the UK but only references to Persons is significant in the USA – and then the effect is negative.

Overall, examining the results for both the USA and UK e-petition platforms it appears that linguistic factors do not significantly impact individually on the popularity of e-petitions. In the UK it would appear that the best strategy to maximise the number of total signatures is to ensure that it is opened during a “quiet” period (where there is little competition from other e-petitions) and that it receives a large number of signatures on its first day of inception. This can be achieved by a successful marketing campaign, including the use of social media. Such an initial burst of signatures is likely to get the e-petition listed as a trending e-petition on the UK Parliament website and noticed by mainstream media, which may garner further signatures. Making the e-petition specific to persons, locations or organisations helps to build the number of signatures, but these signatories may be concentrated in certain spatial locations. Little can be said with regard to topics, with a possibility that a more niche topic area will help to promote the e-petition by minimising its competition with others for signatures. Whilst this article has begun to identify these optimum strategies for increasing the popularity of an e-petition in the UK, the contrasting findings between the USA and the UK suggest that there may not be one universal strategy. To further this understanding, future work looking at other countries may be insightful in demonstrating better alignment, or further interesting differences with the USA or UK. The extension of this study to continental European, South American or Asia countries would also prove insightful, given the potentially larger linguistic and cultural differences from the UK and USA.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.giq.2020.101523>.

Acknowledgements

This work was supported by the Economic and Social Research Council, Project Reference: ES/L011891/1. The authors would like to acknowledge two parties for the supply of the data used in the study. Firstly the data for the UK Parliaments e-petitions were obtained from the Parliament's Petition Committee website archive, and secondly, information on the number of signatures in the first 24 h of an e-petition was supplied by the Oxford Internet Institute. The authors would

also like to thank the reviewers who took the time to consider earlier versions of this article and provide challenging and insightful comments.

References

- Algeo, J. (1986). The two streams: British and American English. *Journal of English Linguistics*, 19(2), 269–284.
- Aragón, P., Sáez-Trumper, D., Redi, M., Hale, S., Gómez, V., & Kaltenbrunner, A. (2018). Online petitioning through data exploration and what we found there: A dataset of petitions from Avaaz. *org. Paper presented at the twelfth international AAAI conference on web and social media*.
- Asai, S. (2009). Sales patterns of hit music in Japan. *Journal of Media Economics*, 22(2), 81–101. <https://doi.org/10.1080/08997760902900106>.
- Barats, C., Dister, A., Gambette, P., Leblanc, J.-M., & Pères, M. (2016). Analyse des pétitions en ligne: potentialités et limites d'un dispositif d'étude pluridisciplinaire. *Paper presented at the JADT 2016*.
- Berger, J., & Milkman, K. L. (2012). What makes online content viral? *Journal of Marketing Research*, 49(2), 192–205. <https://doi.org/10.1509/jmr.10.0353>.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2000). *10.1162/jmlr.2003.3.4-5.993. CrossRef listing of deleted DOIs*. 1(4–5), 993–1022. <https://doi.org/10.1162/jmlr.2003.3.4-5.993>.
- Bochel, C. (2012). Petitions systems: Contributing to representative democracy? *Parliamentary Affairs*, 66(4), 798–815.
- Bochel, C. (2016). Process matters: Petitions systems in Britain's legislatures. *The Journal of Legislative Studies*, 22(3), 368–384.
- Bochel, C., & Bochel, H. (2016). ‘Reaching in?’ The potential for e-petitions in local government in the United Kingdom. *Information, Communication & Society*, 20(5), 683–699. <https://doi.org/10.1080/1369118x.2016.1203455>.
- Briassoulis, H. (2010). Online petitions: New tools of secondary analysis? *Qualitative Research*, 10(6), 715–727. <https://doi.org/10.1177/1468794110380530>.
- Chen, Y., Deng, S., Kwak, D.-H., Elnoshokaty, A., & Wu, J. (2019). A multi-appeal model of persuasion for online petition success: A linguistic cue-based approach. *Journal of the Association for Information Systems*, 20(2), 105–131. <https://doi.org/10.17705/1jais.00530>.
- Clark, S., Lomax, N., & Morris, M. A. (2017). Classification of Westminster Parliamentary constituencies using e-petition data. *EPJ Data Science*, 6(1), 16. <https://doi.org/10.1140/epjds/s13688-017-0113-9>.
- Clark, S. D., Morris, M. A., & Lomax, N. (2018). Estimating the outcome of UKs referendum on EU membership using e-petition data and machine learning algorithms. *Journal of Information Technology & Politics*, 15(4), 344–357. <https://doi.org/10.1080/19331681.2018.1491926>.
- Contamin, J.-G., Léonard, T., & Soubiran, T. (2017). Les transformations des comportements politiques au prisme de l'e-pétitionnement. *Recherches*, 4, 97–131.
- Cream, J., Maguire, D., & Robertson, R. (2018). How have public attitudes to the NHS changed over the past three decades? Retrieved from <https://www.kingsfund.org.uk/publications/how-have-public-attitudes-to-nhs-changed>.
- Davies, C. (2005). *Divided by a common language: A guide to British and American English*. Vol. 13. Houghton Mifflin Harcourt.
- Directorate-General for Internal Policies (2015). The right to petition EU. Retrieved from [http://www.europarl.europa.eu/RegData/etudes/STUD/2015/519223/IPOL_STU\(2015\)519223_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/STUD/2015/519223/IPOL_STU(2015)519223_EN.pdf).
- Dodd, G. (2007). *Justice and grace: Private petitioning and the English parliament in the late Middle Ages*. Oxford: Oxford University Press.
- Dumas, C., LaManna, D., Harrison, T. M., Ravi, S., Hagen, L., Kotfila, C., & Chen, F. (2015a). E-petitioning as collective political action in we the people. *iConference 2015 Proceedings*.
- Dumas, C. L., LaManna, D., Harrison, T. M., Ravi, S. S., Kotfila, C., Gervais, N., ... Chen, F. (2015b). Examining political mobilization of online communities through e-petitioning behavior inWe the people. *Big Data & Society*, 2(2), <https://doi.org/10.1177/2053951715598170>.
- Dunkerley, K. J., & Robinson, W. P. (2016). Similarities and differences in perceptions and evaluations of the communication styles of American and British managers. *Journal of Language and Social Psychology*, 21(4), 393–409. <https://doi.org/10.1177/026192702237956>.
- Egan, S. (2014). National identity and cultural resonance in English foxhunting movements. *National Identities*, 16(1), 71–90.
- Elazar, D. J. (2016). Contrasting unitary and federal systems. *International Political Science Review*, 18(3), 237–251. <https://doi.org/10.1177/019251297018003002>.
- Finkel, J. R., Grenager, T., & Manning, C. (2005). Incorporating non-local information into information extraction systems by Gibbs sampling. *Paper presented at the ACL-05 - 43rd annual meeting of the association for computational linguistics, proceedings of the conference*.
- Fraser, P. (1961). Public petitioning and parliament before 1832. *History*, 46(158), 195–211. <https://doi.org/10.1111/j.1468-229X.1961.tb02441.x>.
- Gil-Garcia, J. R., Pardo, T. A., & Luna-Reyes, L. F. (2018). Policy analytics: Definitions, components, methods, and illustrative examples. *Policy analytics, modelling, and informatics* (pp. 1–16). Springer.
- Gonçalves, P., Araújo, M., Benevenuto, F., & Cha, M. (2013). Comparing and combining sentiment analysis methods. *Paper presented at the proceedings of the first ACM conference on online social networks*.
- Grainger, K., & Mills, S. (2016). *Directness and indirectness across cultures*. Palgrave Macmillan.
- Grun, B., & Hornik, K. (2011). Topicmodels: An R package for fitting topic models. *Journal of Statistical Software*, 40(13), 1–30 Retrieved from < Go to ISI >://

- WOS:000290391200001.
- Hagen, L. (2018). Content analysis of e-petitions with topic modeling: How to train and evaluate LDA models? *Information Processing Management*, 54(6), 1292–1307.
- Hagen, L., Harrison, T. M., & Dumas, C. L. (2018). Data analytics for policy informatics: The case of e-petitioning. *Policy analytics, modelling, and informatics* (pp. 205–224). Springer.
- Hagen, L., Harrison, T. M., Uzuner, Ö., Fake, T., Lamanna, D., & Kotfila, C. (2015). Introducing textual analysis tools for policy informatics: a case study of e-petitions. *Paper presented at the proceedings of the 16th annual international conference on digital government research*.
- Hagen, L., Harrison, T. M., Uzuner, Ö., May, W., Fake, T., & Katragadda, S. (2016). E-petition popularity: Do linguistic and semantic factors matter? *Government Information Quarterly*, 33(4), 783–795.
- Hale, S. A., John, P., Margetts, H., & Yasseri, T. (2018). How digital design shapes political participation: A natural experiment with social information. *PLoS One*, 13(4), Article e0196068. <https://doi.org/10.1371/journal.pone.0196068>.
- Hale, S. A., Margetts, H., & Yasseri, T. (2013). Petition growth and success rates on the UK No. 10 Downing Street website. *Paper presented at the proceedings of the 5th annual ACM web science conference*.
- Hale, S. A., Margetts, H. Z., & Yasseri, T. (2019). *UK parliament petition website*. Hourly count data <https://doi.org/10.1140/epjds/s13688-017-0116-6> Retrieved from:.
- Harrison, T. M., Dumas, C., DePaula, N., Fake, T., May, W., Atrey, A., ... Ravi, S. (2017). E-petitioning and online media: the case of #bringbackourgirls. *Paper presented at the proceedings of the 18th annual international conference on digital government research*.
- Hersh, E. D., & Schaffner, B. F. (2018). Postmaterialist particularism: What petitions can tell us about biases in the policy agenda. *American Politics Research*, 46(3), 434–464.
- Hough, R. (2012). Do legislative petitions systems enhance the relationship between parliament and citizen? *The Journal of Legislative Studies*, 18(3–4), 479–495. <https://doi.org/10.1080/13572334.2012.706057>.
- Houses of Parliament (2017). Petitions: UK government and parliament. Retrieved from <https://petition.parliament.uk/>.
- Jackson, D., Thorsen, E., & Wring, D. (2016). EU referendum analysis 2016: Media, voters and the campaign (1910042099). Retrieved from <http://www.referendumanalysis.eu/>.
- Jongeling, R., Datta, S., & Serebrenik, A. (2015). Choosing your weapons: on sentiment analysis tools for software engineering research. *Paper presented at the 2015 IEEE international conference on software maintenance and evolution (ICSME)*.
- Lee, C.-P., Chen, D.-Y., & Huang, T.-Y. (2013). The interplay between digital and political divides. *Social Science Computer Review*, 32(1), 37–55. <https://doi.org/10.1177/0894439313497470>.
- Leston-Bandeira, C. (2017). What is the point of petitions in British politics? Retrieved from <http://blogs.lse.ac.uk/politicsandpolicy/what-is-the-point-of-petitions/>.
- Manning, C., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S., & McClosky, D. (2014). The Stanford CoreNLP natural language processing toolkit. *Paper presented at the proceedings of 52nd annual meeting of the association for computational linguistics: System demonstrations*.
- Margetts, H., John, P., Hale, S., & Yasseri, T. (2015). *Political turbulence: How social media shape collective action*. Princeton University Press.
- Mullen, T. (2014). The Scottish independence referendum 2014. *Journal of Law and Society*, 41(4), 627–640. <https://doi.org/10.1111/j.1467-6478.2014.00688.x>.
- O'Neil, M., Pineau, M., Kendall-Taylor, N., Volmert, D., & Stevens, A. (2017). Finding a better frame: How to create more effective messages on homelessness in the United Kingdom. Retrieved from https://www.crisis.org.uk/media/237700/finding_a_better_frame_2017.pdf.
- Porshnev, A. (2018). Success factors of electronic petitions at Russian public initiative project: The role of informativeness, topic and lexical information. *International Conference on Social Informatics (Vol. Lecture Notes in Computer Science)*. 11186. *International Conference on Social Informatics (Vol. Lecture Notes in Computer Science)* (pp. 243–250). Cham: Springer International Publishing.
- Potts, A., & Baker, P. (2012). Does semantic tagging identify cultural change in British and American English? *International Journal of Corpus Linguistics*, 17(3), 295–324.
- Puschmann, C., Bastos, M. T., & Schmidt, J.-H. (2016). Birds of a feather petition together? Characterizing e-petitioning through the lens of platform data. *Information, Communication & Society*, 20(2), 203–220. <https://doi.org/10.1080/1369118x.2016.1162828>.
- R Core Team (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Salathé, M., Vu, D. Q., Khandelwal, S., & Hunter, D. R. (2013). The dynamics of health behavior sentiments on a large online social network. *EPJ Data Science*, 2(1), 4. <https://doi.org/10.1140/epjds16>.
- Sheppard, J. (2015). Online petitions in Australia: Information, opportunity and gender. *Australian Journal of Political Science*, 50(3), 480–495. <https://doi.org/10.1080/10361146.2015.1049512>.
- Slawson, N. (2016). *Brexit: No second EU referendum despite e-petition, says government*. The Guardian. Retrieved from <https://www.theguardian.com/politics/2016/jul/09/government-dashes-hopes-of-second-eu-referendum-in-e-petition-response>.
- Slawson, N. (2017). *Anti-trump petition to stop UK state visit passes 1m signatures*. The Guardian. Retrieved from <https://www.theguardian.com/us-news/2017/jan/30/anti-trump-petition-to-prevent-uk-state-visit-passes-1m-signatures>.
- Socher, R., Perelygin, A., Wu, J. Y., Chuang, J., Manning, C. D., Ng, A. Y., & Potts, C. (2013). Recursive deep models for semantic compositionality over a sentiment treebank. *Paper presented at the proceedings of the conference on empirical methods in natural language processing (EMNLP)*.
- Sorensen, A. T. (2004). Bestseller lists and product variety: The case of book sales. Working paper no. 1878. Retrieved from <https://www.gsb.stanford.edu/faculty-research/working-papers/bestseller-lists-product-variety-case-book-sales>.
- Suh, J. H., Park, C. H., & Jeon, S. H. (2010). Applying text and data mining techniques to forecasting the trend of petitions filed to e-people. *Expert Systems with Applications*, 37(10), 7255–7268. <https://doi.org/10.1016/j.eswa.2010.04.002>.
- Taagepera, R. (1972). The size of national assemblies. *Social Science Research*, 1(4), 385–401.
- TeBlunthuis, N. (2018). Density dependence without resource partitioning on an online petitioning platform. Retrieved from https://digital.lib.washington.edu/researchworks/bitstream/handle/1773/41690/TeBlunthuis_washington_02500_18210.pdf.
- Vidgen, B., & Yasseri, T. (2020). What, When and Where of petitions submitted to the UK Government during a time of chaos. *Policy Sciences*, 53(3), 535–557.
- Wellings, D. (2018). What does the public think about the NHS? Retrieved from <https://www.kingsfund.org.uk/publications/how-have-public-attitudes-to-nhs-changed>.
- Wright, S. (2015a). E-petitions. In S. Coleman, & D. Freelon (Eds.). *Handbook of digital politics* (pp. 29). Cheltenham: Edward Elgar Publishing Limited.
- Wright, S. (2015b). “Success” and online political participation: The case of downing street E-petitions. *Information, Communication & Society*, 19(6), 843–857. <https://doi.org/10.1080/1369118x.2015.1080285>.
- Yasseri, T., Hale, S. A., & Margetts, H. Z. (2017). Rapid rise and decay in petition signing. *EPJ Data Science*, 6(1), 13. <https://doi.org/10.1140/epjds/s13688-017-0116-6>.
- Anthony, E and Haworth J (2020). A spatial analysis of UK e-petition signing behaviour. Paper 73, GIS Research UK Conference 2020, July, London, United Kingdom. http://london.gisruk.org/gisruk2020_proceedings/GISRUUK2020_paper_73.pdf (Accessed: 15/09/2020).

Stephen Clark is a Research Fellow in the Leeds Institute for Data Analytics and the School of Geography, University of Leeds. His research interests include the application of innovative data analysis techniques to novel data from the domains of housing, transport, politics and lifestyles.

Nik Lomax is an Associate Professor in Data Analytics for Population Research at the University of Leeds. His research utilises existing and emerging data to produce high resolution estimates and projections of populations and their demographic characteristics. Nik is also a Turing Fellow, where he is developing methods to integrate information on human behaviour which are lacking from current demographic models.