Mining public opinion: why unsuccessful online petitions should not be ignored

Taha Yasseri *argues that by analysing online petition data using computational techniques, politicians can glean fresh insights about the geographic factors influencing constituents' concerns, the dynamics at play over time, as well as a deeper awareness of the issues most important to the general public.*

Petitions are an excellent data source for understanding the concerns and priorities of citizens. They can be considered 'big data' as they contain large amounts of time-stamped granular transactional data and are available in real-time. However, they are under-utilized in social scientific research and government services. The focus has been limited to the most popular petitions. The remaining petitions which fail to receive enough signatures will turn into 'digital dust'. Ironically, this latter group of petitions can make up to 99% of all the petitions that are approved to appear on the petitioning website.

In the current unpredictable and chaotic political environment, there is an even greater need for governments to understand the concerns of the public, and to reflect these in their agenda, discourse, and policies. Signing a petition is one of the few ways in which citizens can easily and legally raise issues in between elections. In a paper co-authored with Bertie Vidgen, we computationally analysed all petitions submitted to the UK government between the 2015 and 2017 general elections.

We used unsupervised machine learning algorithms to extract petitions' 'topics'. Instead of reading and manually coding the 11,000 petitions under study, we used a systematic computational approach in analysing the content of petitions known as 'topic modelling'. We extracted ten most tightly packed groups of words that are likely to appear together in a single petition. We considered these bags of words as the topics and manually annotated them into ten issues. We then assigned each petition to one of these ten topics computationally, based on the similarity between their content and the words pre-assigned to topics.

Based on the number of signatures, we find that the most prevalent issue is 'Democracy & the EU' (7.5 million signatures), followed by 'International Affairs' (5.8 million signatures) and 'Healthcare' (3.1 million signatures).

Figure 1: the prevalence and success of issues included in online petition
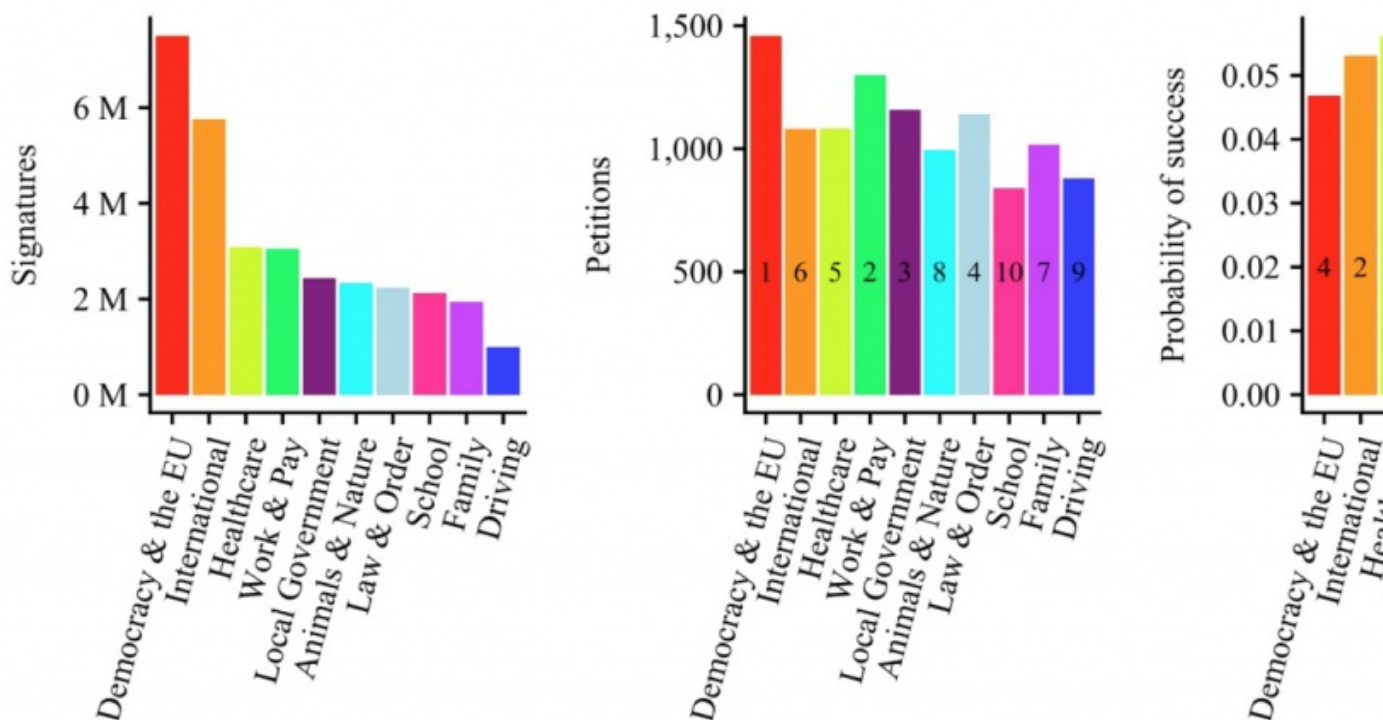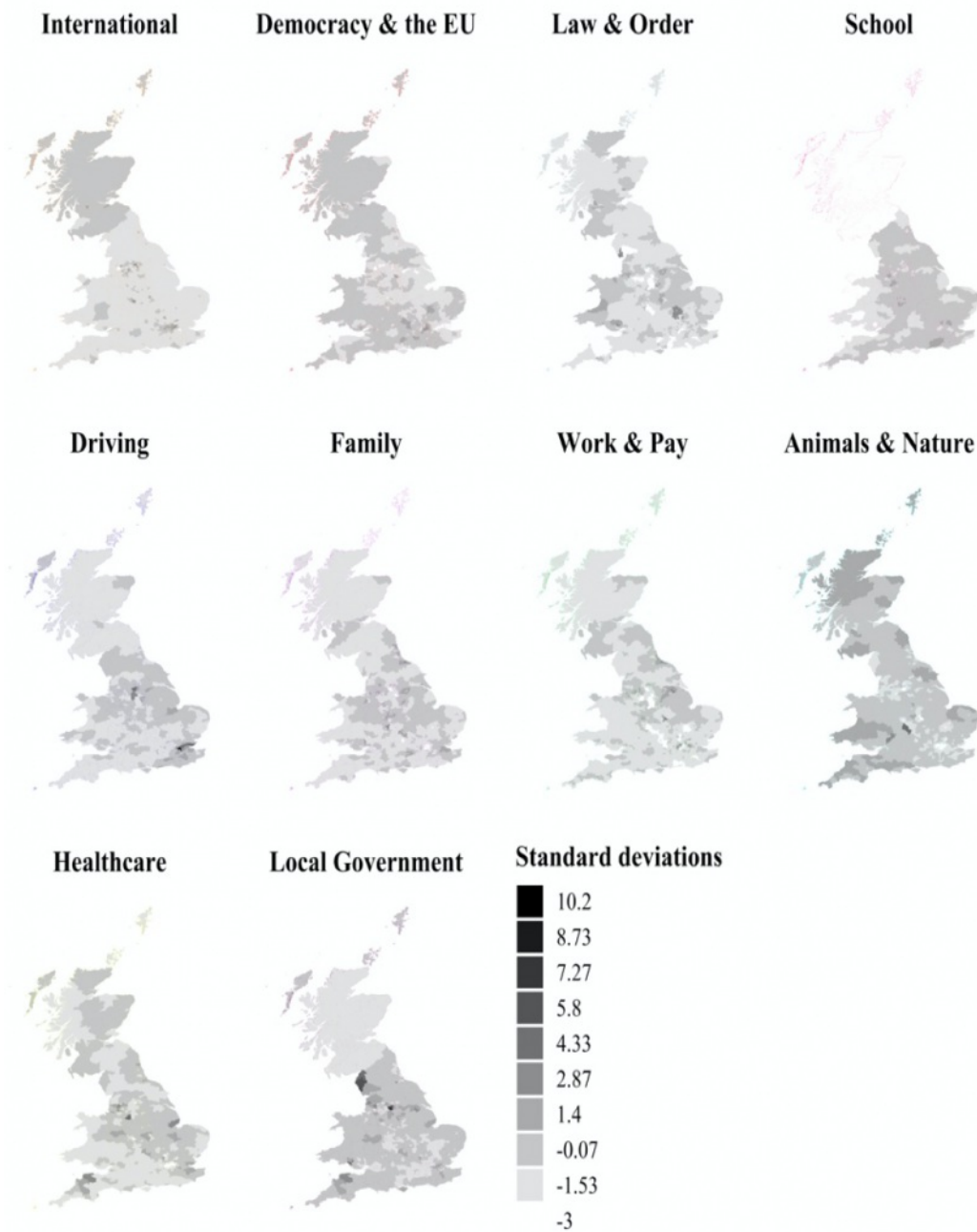

Left: the distribution of signatures over issues. Middle: the distribution of petitions over issues. Right: the probability that petitions assigned to each issue will receive 10,000 signatures or more. The numbers show the issues' ranked position. Adopted from Vidgen, B., Yasseri, T.

Issues show different temporal dynamics, whereby some exhibit large fluctuations over time and others exhibit minor fluctuations in prevalence and popularity. Our analysis shows a very noticeable spike for the issue 'Democracy & the EU' in May 2016, due to a popular petition which called for the EU referendum vote to be repeated. Similarly, the issue 'International Affairs' has large spikes in November 2016 due to a highly-publicized petition which called for Donald Trump to be banned from visiting the UK. This suggests that for both these issues, signatures were driven by exogenous events. In contrast, for the other eight issues we analysed, fluctuations decrease noticeably as the time window increases, which suggests that signatures are broadly stable and not driven by external events. Monitoring the temporal dynamics of issue popularity can be considered as a simple yet effective approach in gauging volatility and diversity of the public discourse.

Our analysis also shows that different geographic areas sign petitions associated with different topics. We find that several issues can be identified as national issues, including, 'Law & Order' and 'Work & Pay', with the geography of signatures given to them being more uniform than for other issues. They attract support from many different parts of the country, and the variations do not follow a discernible pattern.

In contrast, other issues are highly regional. For example, 'Driving' is highly important for a small set of constituencies in the South East but less so elsewhere. 'Animals & Nature' is also particularly notable: urban areas, including London, the Midlands, and northern cities assign very little attention to the issue; rural constituencies assign more attention; and areas of natural beauty which are far from urban centres, including Cornwall, West Wales and North Scotland, assign it the most importance. This reflects a broader pattern where, in general, petition-signing habits vary between rural and urban constituencies. Rural constituencies tend to petition about traditional domestic political issues whilst urban areas are more concerned about ideological issues.

Figure 2: prevalence of issues covered in online petitions by area.

International   Democracy & the EU   Law & Order   School

Driving   Family   Work & Pay   Animals & Nature

Healthcare   Local Government   Standard deviations

	10.2
	8.73
	7.27
	5.8
	4.33
	2.87
	1.4
	-0.07
	-1.53
	-3

The prevalence of issues in each constituency. The darkness of the shading represents the number of standard deviations the percentage of signatures from each constituency for each issue are from the mean. Adopted from Vidgen, B., Yasseri, T.

We then went one step further, by investigating the relationship between geography and petitions and identifying clusters showing distinct regions of petition signing, which complemented our earlier findings. Our clustering analysis shows a clear issue divide between rural and urban constituencies, and sets out a distinctive region which is mostly comprised of only Scotland.

Our analysis indicates that the concerns of citizens as expressed through petitions are linked powerfully to not only temporality, and the impact of exogenous events, but also geography. This opens up new avenues for research, including investigations of how geographic environment influences individuals' behaviour and how geography can be a proxy measure for other issue-influencing factors, such as ethnicity, class and gender. Following Margetts and Dorobantu's suggestion to 'rethink the government with AI', our research demonstrates how the thematic content of petitions can be analysed by machine learning methods in order to understand the issues which concern the public. It also shows that the UK public's interest in issues is complex and heterogeneous: there are important geographic and temporal dynamics which should be taken into account by decision-makers.

In the long term, there is also scope for integrating analysis of petitions' content with their sentiment and ideological stance. By doing so, politicians could benefit from even deeper real-time insights into how their constituents view key issues and, most importantly, how they want those issues to be addressed (taking into account, of course, the digital divide and the uneven rate of participation in online initiatives).

_____

Note: the above draws on the author's published work (with Dr Bertie Vidgen) in Policy Sciences journal.

About the Author

Taha Yasseri is Associate Professor of Sociology at University College Dublin, a Former Senior Research Fellow in Computational Social Science at the Oxford Internet Institute, University of Oxford, and a Turing Fellow at the Alan Turing Institute for Data Science.

Photo by Florian Olivo on Unsplash