Paper to the Annual Conference of the UK Political Studies Association
Cardiff Wales
25-27 March, 2013

Understanding the Dynamics of Internet-based Collective Action using Big Data: Analysing the Growth Rates of Internet-based Petitions

Scott Hale, Helen Margetts, Taha Yasseri
Oxford Internet Institute, University of Oxford
{scott.hale,helen.margetts, taha.yasseri}@oii.ox.ac.uk

Abstract

Now that so much of collective action takes place on Internet-based platforms, it leaves a digital imprint which may be harvested to further understanding of the dynamics of mobilization. This 'big data' offers social science researchers the potential for new forms of analysis, using real-time transactional data based on entire populations, rather than sample-based surveys of what people think they did or might do. This paper uses a big data approach to track the growth of over 30,000 petitions to the UK Government on two platforms over two years, analysing the rate of growth and testing the hypothesis that the distribution of daily change will be leptokurtic (rather than normal) as previous research on agenda setting would suggest. This hypothesis is confirmed, suggesting that Internet-based mobilization is characterized by tipping points (or punctuations) and explaining some of the volatility in online collective action. We find that the vast majority of petitions fail to attain even minimal levels of success. Where there is success, the punctuations occur very early in the life of a petition. We propose a model of 'collective attention' to characterise this rapid growth and early saturation. These findings have implications for the strategies of those initiating petitions and the design of web sites with the aim of maximising citizen engagement with policy issues.

Keywords **Collective Action, Big Data, Petitions, Attention**

Introduction

Now that so much of collective action takes place on Internet-based platforms, most mobilizations include a digital element and some take place almost wholly online. All this Internet-based activity leaves a digital footprint, which allows researchers to generate 'big data', real-time transactional data of political behaviour. This kind of data presents new opportunities for social science research, allowing a move away from sample-based survey data about what people think they did or will do, to transactional data about what people actually did based on whole populations. Big data is receiving massive attention and interest across the corporate world and scientific research communities (Mayer Schoenberger and Cukier, 2013). Yet the potential of this kind of data for understanding political behaviour remains under-explored. We are still at the start of exploiting the great potential of 'big data' to explore the dynamics of Internet-based collective action and how it may challenge existing theories and models of collective behaviour.

This paper analyses two 'big data' sets of Internet-based mobilizations, generated from two electronic petition platforms; first, the e-petition part of the No.10 Downing Street web site (from February 2009 to the point at which it closed in March 2011) and second, the new e-petition site developed by the UK Cabinet Office for the incoming Coalition Government in 2010 and launched in August 2011, which replaced it. E-petitions were first launched in the UK in November 2006, and over the course of its lifetime the No. 10 site received more than 8 million signatures from over 5 million unique email addresses.[1] Both sites have allowed anyone to view petitions, and any user with a valid email address could create a new petition or sign an existing petition. There are important differences between the sites however; for example, whereas the first site showed the names of the 500 most recent petitioners, the new site shows only the name of the petition creator. The sites have also provided alternative measures of 'success' of a petition; for example, for the earlier site, the government promised an official response to all petitions receiving at least 500 signatures, while the incoming Coalition government in 2010 promised that any petition on the new site attracting more than 100,000 signatures would qualify for a parliamentary debate on the issue raised.

This paper analyses the growth of petitions and the distinctive characteristics of the mobilization curves of successful and unsuccessful petitions on both platforms. First, we provide some background on online collective action in general and e-petitions in

1 http://www.mysociety.org/projects/no10-petitions-website/

particular. We use previous work on political attention to develop a hypothesis regarding the development of online mobilizations: that they will be characterised by long periods of stasis and short periods of rapid change, leading to a leptokurtic distribution of daily change. Second, we outline the methods used to test this hypothesis and third, provide the results for two big data sets, harvested from two e-petition platforms provided by the UK government, followed by a discussion of the policy and design implications of the findings.

Background

The 21st century has seen a prominent role for the Internet in mobilization, from the dramatic events in authoritarian states of the Arab Spring, to a series of protests, demonstrations and social backlash against austerity driven cutbacks and state retrenchment in liberal democracies facing the consequences of the financial crash of 2008. Researchers have turned their attention to the theoretical and conceptual implications of online collective action (e.g. Bennett and Segerberg, 2012; Bimber, 2003; Lupia and Sin, 2005) and some are using innovative methods, including experiments and data-mining, to explore the viral spread of mobilizations across online social networks (see, for example, Ackland, 2007; Etling et al, 2009; Hindman, 2008; Gonzalez Bailon et al, 2011; Segerman and Bennett, 2011; Aral and Walker, 2010).

While online activity may be a minor element of some mobilizations, other mobilizations occur almost entirely online. The trend for 'e-petitioning' represents one such activity, where online petitions are created, disseminated, circulated, and presented online, and although policy-makers may discuss responses in offline contexts, such responses are generated and sent online. The UK government's e-petition site was created by the social enterprise MySociety on the No. 10 Downing Street website in November 2006 and ran until March 2011, when it was closed by the incoming Coalition government. Some of these petitions had high policy impact, notably one against the Labour administration's proposed road pricing policy, which policy-makers admitted off the record played a role in getting the policy scrapped. A new site was launched in August 2011 by the Cabinet Office, initially on the direct.gov portal, which transmuted into the new portal www.gov.uk in the autumn of 2012, with a different format. Signing petitions has long been among the more popular political activities, leading the field for participatory acts outside voting and with other social benefits ascribed to it as well as

having the potential to bring about policy change; e-petitioning reinforces 'civic mindedness' (Whyte et al. 2005) and is one of a growing portfolio of Internet-based democratic innovations (Smith, 2009). The widespread use of e-petitions by both governments and NGOs such as Avaaz and 38 Degrees has received accolades for their democratic contribution (Escher, 2011; Chadwick, 2012) and the German e-petition platforms have been analysed extensively (see Lindner and Riehm, 2011; Jungherr and Jurgens, 2010), but the UK petition platforms have received rather less attention in recent political science research, with the exception of Wright (2012).

E-petitions are interesting examples of mobilizations with a strong online imprint, which will include the entire transaction history for both successful and unsuccessful mobilizations. The data that can be harvested from the signing of electronic petitions is an example of what is now commonly known as 'big data', representing a transactional audit trail of what people actually did (as opposed to what people think they did) and an entire population (without the need to take a representative sample). Data like this represents a big shift for social science research into political behaviour, which has traditionally rested on survey data, or, for elections, voting data. Big data also presents challenges to social science research—it doesn't come with handy demographics attached and we do not know where people came from to any one interaction, nor where they are going, so it is often difficult to match up online activities across different platforms, or to identify the underlying factors influencing behaviour, such as age, income or gender. This data however, makes it possible to look at the different patterns of growth in the 30,000 mobilization curves that we have and identify the distinctive characteristic of those mobilizations that succeed and those that fail with our digital hindsight. Such an analysis, using data that has rarely been available to political science researchers before the current decade, may tell us something about the nature of collective action itself in a digital world. Of the research noted above, Jungherr and Jurgens used a smaller dataset to illustrate the viability of a big data (or computational social science) approach, but other studies used surveys (Lindner and Riehm, 2011) or more qualitative approaches (Wright, 2012).

So what would we hypothesise about these mobilizations? A possible hypothesis may be derived from previous research on agenda setting in political systems. The most well known model of how policy attention proceeds in a liberal democracy is that of 'punctuated equilibrium', developed by the US authors Baumgartner and Jones and their 'Policy Agendas' programme of research (see www.policyagendas.org). The theory

argues that policy attention to any issue will remain in long periods of stasis where little change occurs. Where issues do hit the policy agenda, it will be because some event has 'punctuated' the equilibria, all eyes (including the media, public opinion, interest groups and politicians concerned) turn to the issue, money is spent, institutions are created and policy change occurs (John and Margetts, 2003; Baumgartner and Jones, 1993; Jones and Baumgartner, 2005). The theory of punctuated equilibria is multi-faceted and has been illustrated by a range of empirical data across policy areas and within different dimensions of attention, such as public opinion, budgetary change and congressional attention (Baumgartner and Jones, 2005) and in various countries, including the UK (John and Margetts, 2003). Baumgartner and Jones do not discuss Internet-related activity to any great degree; however, we might hypothesise that the pattern of mobilizations around a petition would proceed in a similar way, thereby contributing to the same sort of issue attention cycle that has been observed many times over in agenda setting research. Such a model would predict that the distribution of daily changes in attention would be 'leptokurtic', with a small number of large changes and a longer tail of much smaller ones.

Such a finding could not show any causal effect, as only the activity of petitioners is being analysed here (and tipping points could suggest a media effect, although it is extremely unlikely that this could take place before substantive amounts of signatures had amassed). It could, however, point towards a role for online mobilization in policy change analogous to that of the media in the agenda setting analysis, which is ascribed a lurching effect, due to the capacity of the media to parallel process only a small number of issues; at the point at which a punctuation occurs, media attention will 'tip over' from specialist outlets into the mainstream media. In addition, in lower level mobilizations where the media is not paying attention, such a finding could suggest a role in policy issues below the media radar.

Methods

The UK Government's petition website (petitions.number10.gov.uk) was accessed daily from 2 February 2009 until March 2011, when the site closed, with an automated script. Each day, the number of overall signatures to date on each active petition was recorded. In addition, the name of the petition, the text of the petition, the launch date of the petition, and the category of the petition were recorded. Overall, 8,326 unique petitions

were tracked from the earlier site, representing all publically available petitions active at any point during the study. A first pass at this data after the site closed revealed the importance of the first day in the future of a petition (see below and Hale and Margetts, 2012) and suggested that more frequent scraping of the data could deliver a more fine-grained analysis. For this reason, when the new site was launched in August 2011, we set the automatic script to scrape it every hour, recording the same details as for the previous site. Our second dataset currently contains hourly data points for all the petitions (19,789) submitted to the new site from 5th August 2011 to 22nd February 2013.

The two sites differed under different policies for how the government would respond to petitions. For the No 10 Downing Street site, prospective petitioners were told that if their petition achieved 500 signatures, they would receive an official response. There were no other official measures of success, although one petition did succeed in raising over one million signatures, which previous research has identified as a possible 'tipping point' for mobilizations; if potential participants know that more than one million have participated, they are more likely to participate themselves (Margetts et al, 2011); the petition was widely regarded as influential in getting the policy reversed. For the Cabinet Office site, the bar for an official response is unclear from the site, although the majority of petitions that have over 10,000 signatures do receive a response with the prefix 'As this e-petition has received more than 10,000 signatures, the relevant Government department have provided the following response'. More importantly, in the early days of the Coalition Administration, David Cameron promised that signatures obtaining more than 100,000 signatures would generate a parliamentary debate on the issue raised by the petition. All these information cues will have acted as possible drivers on individuals considering whether to sign a petition.

Petitions on the Number 10 website closed, by default, 12 months after they first launched. To identify patterns in how petitions grow, the percentage change in new signatures was calculated each day, for the earlier site. Most petitions had a long period of inactivity prior to their deadline date. To consider just how petitions grow, data was truncated after the last signature on a petition, removing any final period of zero signature-per-day growth prior to the petition's deadline.

Leptokurtic distributions have a more acute peak close to the mean and larger tails. There is no statistical test to specifically classify a distribution as leptokurtic. However, several tests in combination help demonstrate a distribution is leptokurtic (see John and Margetts, 2003). The most rigorous test is the Shapiro-Wilk test (1965), which checks

whether the points could possibly be drawn randomly from a normal distribution. Leptokurtic distributions should reject the Shapiro-Wilk null hypothesis of normality. The Kolmogorov-Smirnov test (Chakravarti et al. 1967, pp. 392-4) tests that a set of frequencies is normal distributed by focusing on the skewedness and kurtosis of a distribution, and this null hypothesis should be rejected if a distribution is leptokurtic and hence non-normal. Visualizing the histogram and plotting a log-lot plot, which should be nearly a straight line if changes are leptokurtic, give further evidence of a leptokurtic distribution.

Results

First, we explored the data harvested from the first e-petitions site on the No. 10 Downing Street web site, which produced a set of 8,326 unique petitions, shown in Figure 1. The most immediate finding of interest was that 94 per cent failed to obtain even the modest 500 signatures required to elicit an official response, the only measurable 'success' indicator for the earlier site. Nearly all petitions that succeeded in obtaining 500 signatures did so quickly. Successful petitions took a mean time of 8.4 days to reach 500 signatures, but a median time of only two days. In fact, 230 of the 533 successful petitions succeeded in obtaining 500 signatures on the day they were launched (day 1). Only a few petitions take a much longer time to reach the 500 signature mark: 31 petitions (6 per cent) succeed after taking more than 30 days, and only five petitions in our dataset reached the 500 signature mark after being active more than four months.

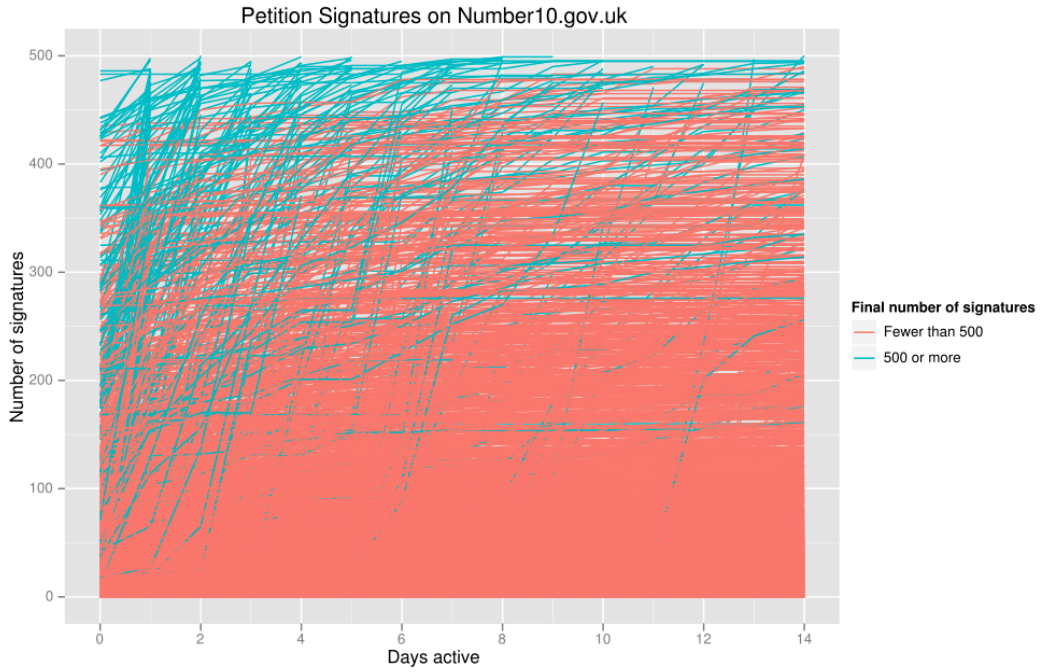
Next, we tested our hypothesis that the distribution of daily change in signatures would be leptokurtic. Figure 2 shows the percentage change in new signatures adjusted so that the mean growth of each petition lies at zero. While most daily change is small, petitions' growth is punctuated by a few large changes. The distribution of growth is leptokurtic and strongly rejects the Shapiro-Wilk null hypothesis of normality with a w statistic of 0.17 translating to a p-value less than 0.000001. The distribution has a kurtosis score of 1,445 and a skewedness of 30.53, and rejects the Kolmogorov-Smirnov test for a normal distribution ($p<0.0001$). When we applied the same tests to the population of petitions that were successful in achieving 500 signatures (that is, excluding the unsuccessful ones), we found a similar leptokurtic distribution (Shapiro-Wilk w statistic of 0.10, $p<0.000001$).

So having identified punctuations, what can we say about where they are? The

largest daily changes happened at the start in the life cycle of the petition. Looking at the distribution of the day on which the punctuation occurred, we see that (for example) all daily changes of more than 80 per cent occurred in the first five days, and greater than 40 per cent in the first 8 days 0 and even for all changes over 40 per cent, the median day is 1, and the mean is 2.2 and the third quartile is one.

So it seems that the early days of a petition are crucial, in particular the first day. Running a logit regression revealed that the number of signatures a petition received on its first day is the most important factor in explaining the petition's success, and a linear regression (shown in Table 1) showed that it was also the most important factor in explaining the total number of signatures the petition receives during its lifetime. All other factors tested—the topic category, the start day of the week, weekend vs weekday launch—did not have a significant effect on the growth of a petition once controlling for the number of signatures the petition received on its first day. Petitions tended to grow shortly after launch and then stop growing. This active period of growth for petitions has a mean length of 57 days and a median length of 27 days.

Figure 1. Petition Growth on No. 10 Downing Street e-petitions site, 2009 – 2010



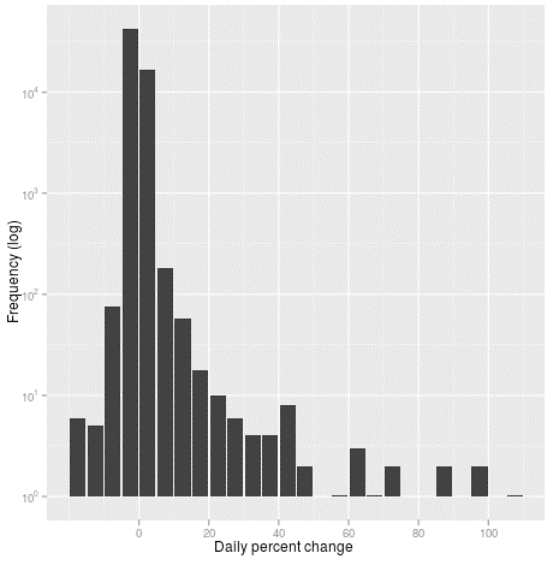
Note: N = 8,326 petitions, all petitions created between September 2009 and May 2010

Table 1: Factors affecting growth. OLS predicting the total number of signatures

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	218.41	78.55	2.78	0.00544 **
Category	-6.37	4.97	-1.28	0.19976
Day of week petition started	-12.23	11.79	-1.04	0.29974
Signatures collected on first day	1.95	0.02	96.29	<0.00001 ***
Number of petitions started on the same day	-1.55	2.96	-0.53	0.59960

Note: Adjusted R-squared: 0.5268

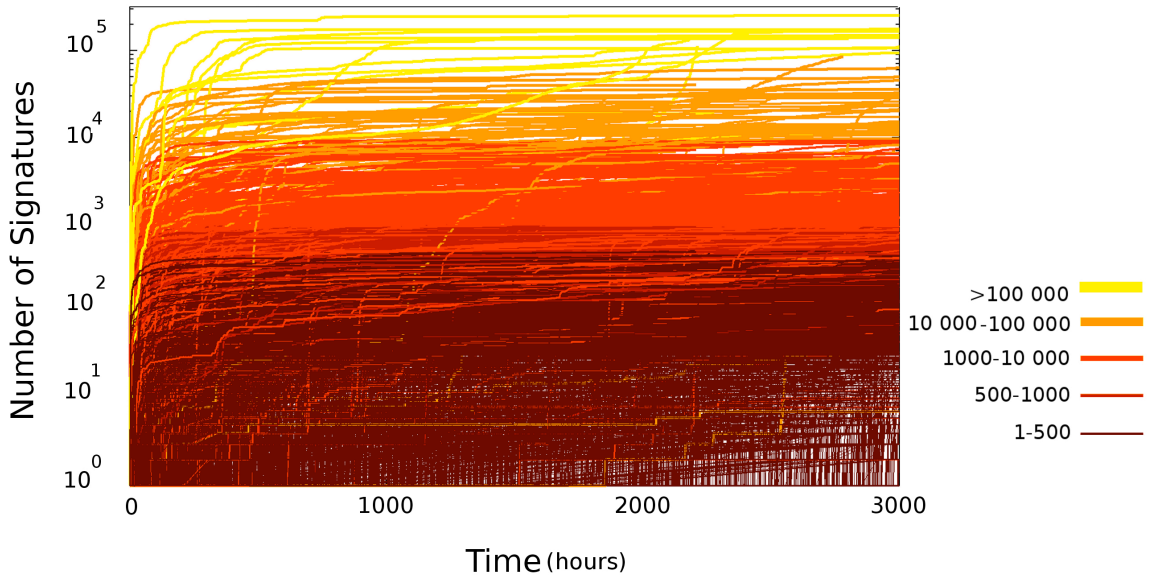
Figure 2: Log of daily percentage change in number of signatories (centred around each petition's mean).



Note: Any final period in which petitions that gain no more signatures after a certain point before closing date has been removed from the daily per cent change data (so the tallest bars do not include these 'zero change' days.

The identification of the first day as a key 'tipping point' for successful petitions caused us to re-examine our data collection techniques as we took the data harvesting and analysis forward with the new site that opened in the summer of 2011. That is, the mobilization on this first day was so rapid that daily scraping of the site could not provide a fine-grained analysis of this or other potential tipping points. Therefore, as noted above, when we collected data from the new site we did so every hour, providing us with 19,789 petitions up to February 2013 (we continue to collect the data). The analysis of this dataset in this draft paper, however, uses a randomly selected subset of this new data. For this subset, Figure 3 shows (in different colours) those petitions that attained the first level of 'success' (that is, the 10,000 signatures required for an official response); and those that attained the second level, the 100,000 signatures required to generate a parliamentary debate.

Figure 3: Petition growth on Cabinet Office e-petitions site, 2011 - 2013



Note: Graph shows a fraction of N=3813 petitions, all created between 5th August 2011 and 22nd February 2013. Note also that y-axis uses a logarithmic scale.

Once again, it is immediately clear that the vast majority of petitions did not achieve any

measure of success. Only 5 per cent of petitions obtained the 500 signatures, which we calculated to compare with the previous dataset, and only 4 per cent received 1,000. Only 0.7 per cent attained the 10,000 signatures which seems to be the bar for receiving some sort of official response, and only 0.1 per cent attained the 100,000 required for a parliamentary debate. Again, the first day was crucial to achieving any kind of success. Any petition receiving 100,000 signatures after three months, needed to have obtained 3,000 within the first 10 hours on average.

Discussion

The results from both data sets show just how few petitions actually attain success by any measure. For the earlier data from the No.10 Downing Street site, the 500 signature mark seems at first consideration a very low threshold that should easily be passed. However, by far the majority of petitions (94 per cent in this time period) fail to attain even this modest number of signatures, illustrating the point that in online environments, the low costs of initiating a collective action mean that there are likely to be large numbers of unsuccessful mobilizations. Petitions are most active when they are first launched and most petitions (presumably in the lack of outside stimulus) become digital dust after a couple of months despite typical deadlines of one year on the site. Our second data set tells the same story, suggesting that this finding may be generalized to other mobilizations rather than representing some characteristics of the No. 10 Downing Street platform. The finding that most mobilizations of this kind fail to take off in any sense chimes well with recent research into the spread or diffusion of initiatives across online networks. Goel et al (2012), for example analyse the diffusion patterns arising from online domains, ranging form networked games to microblogging services and find that in all their seven cases, the vast majority of cascades are small, and are described by a handful of simple tree structures that terminate within one degree of an initial adopting 'seed'. Even for the few large cascades that they observed, the bulk of adoptions often take place within one degree of a few dominant individuals. Although we have not yet made an attempt to model the network activity behind the petitions studied here, it seems likely that to do so would reveal a similar pattern.

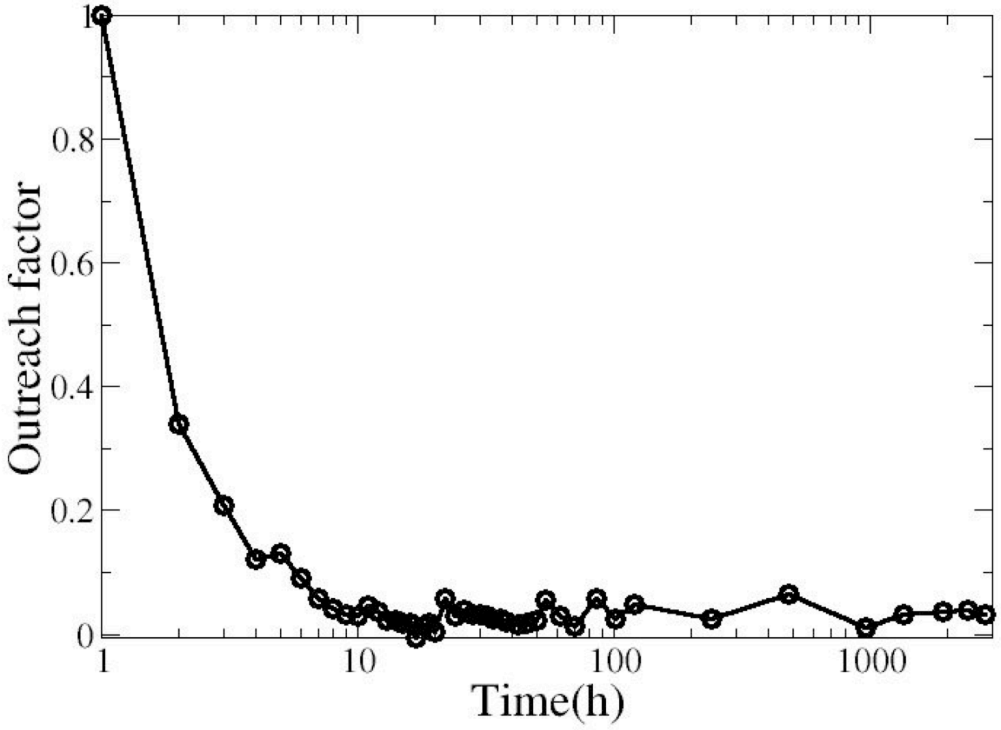
We attempt to capture the characteristic of early rapid growth and decay that the data reveals, with a model of 'collective attention' decay, drawing on Wu and Huberman (2007). In their model, they calculate a 'novelty' parameter relating to the novelty of news items on a news sharing platform, but in a more general framework, the decay in

attention could have other reasons, for example reaching the system size limits, or lack of viral spread. In the model, N agents at the time t, bring $N\mu$ new agents in the next step in average, μ being a multiplication factor; in our case, this would mean that every signature brings μ new signatures in the next hour, leading to an exponential growth of rate μ in the number of signatures. This model works quite well at the beginning, but very soon the spread rate decays and new signatures come at a much lower rate. Therefore we let the multiplication factor decay by introducing a second factor *r(t),* which decays in a way that is an intrinsic of the medium; each signature at time t, on average brings $\mu r(t)$ new signatures in the next hour. To correct for the early saturation observed in the empirical results, we enter an 'outreach' parameter which can change over time and damp the fast initial growth. To calculate that empirically, we average over the logarithm of the number of signatures in hourly bins, starting from the launching time and then calculate the hourly increment at time t and normalize it by the logarithm of the number of signatures up to time t as follows:

$$r(t) = \frac{E[log(N(t))] - E[log(N(t-1))]}{E[log(N(t))]} \qquad (1)$$

In other words, the outreach factor measures the relative growth of the logarithm of number of signatures within an hour, averaged over the whole sample. This parameter is shown as a function of adjusted time in Figure 4. This shows that collective attention decays very fast indeed, and that after 24 hours, a petition's fate is virtually set.

Figure 4. Rate of Change of 'Collective Attention' paid to Petitions created on the Cabinet Office site calculated based on Eq. 1.



Confirmation of our hypothesis regarding the leptokurtic distribution of changes to the support for a petition suggests that in online environments, collective action could play a role in a punctuated equilibrium model of policy change. That is, the general pattern for policy attention is for issues to remain dormant or in stasis, with a generally low level of attention. Some issues (by far the minority) that attract attention quickly gain a 'critical mass' of activists and start to vie for policy attention, joining the range of other institutional influences in helping to 'punctuate' the equilibrium. Such an argument would not include the claim that the mechanism by which collective action acts to bring about instability would be the same as the role played by the media, which plays a distinctive 'lurching' role in Jones and Baumgartner's analysis, based in part on the tendency of the media to process a small number of issues in parallel. In the context studied here, the mechanism would depend more on the ways in which a mobilization is disseminated via online social networks, something that previous research mentioned in the introduction has begun to investigate. If such activity tends to take place on the day

the petition is initiated, then these findings could indicate the importance of achieving some kind of viral spread across the petitioner's closest contacts right at the start, because initial rapid growth will have a greater effects on subsequent participation decisions by 'weaker tie' contacts than a gradual growth over a prolonged period.

In other empirical studies using experimental methodologies, we have started to uncover the mechanism behind such punctuations. That is, in addition to the close contacts informed about the petition directly by the petitioner, other individuals (including people unknown to the petitioner or more distant contacts) deciding whether to participate will be influenced by the information that other people have already participated. This influence will depend on the number of other participants (Margetts et al, 2011), the personality of the individual deciding whether to participate (Margetts et al, 2013), and the closeness of the individual to the petitioner. The occurrence of a punctuation will depend on the existence of 'starters' whose thresholds for participation are low or whose closeness to the petitioner has in this instance reduced their threshold for participation. These starters will act as a signal for people with higher thresholds and weaker ties to the petitioner to 'follow' in signing the petition, thereby acting as a further signal for people with even higher thresholds to join. At some point, if the petition is successful, then the number of followers will reach 'critical mass' and attention to the mobilization will become widespread, breaking out of the petitioner's social network and gain more general social media exposure.

So, there are two possible explanations for the importance of the first day in achieving the critical mass: first, the way in which the petition is disseminated via online networks and second, the dynamic of starters and followers and social information about other participants in the burst of activity. If it is that visitors to the petition are heavily influenced by the numbers of other participants that they observe, then the finding could join those of other work in economics illustrating the importance of 'first donations' in charitable giving (Bog et al, 2006), showing that early donors set the precedent for later donors, or the wider literature on conditional co-operation, showing that social information about the contribution of others influences an individual participant's decision to contribute (Croson and Shang, 2009; Frey and Meier, 2004). This study complements this previous work by focusing on numbers of participants, rather than contribution amount (as everyone's contribution, at least that we are able to measure, is the same) and a more explicitly political context, rather than that of charitable giving.

The strong effect of petitions tending to succeed quickly or not at all will be

influenced by the design of the petition website during the period of study; the 'outreach' factor will vary across platforms. For users starting at the homepage of the earlier No.10 Downing St site, it was possible to view petitions overall or within a specific category and to sort petitions by the number of signatures or the date added. It was therefore easiest to look at petitions with the largest or smallest number of signatures and the oldest or newest petitions. On the newer Cabinet Office site, petitions can be sorted by signature or closing date, or viewed by government department, but not by topic. In addition, we can expect different behaviour from users of either site who arrive at the homepage (who may respond to these information cues by looking only at the newest petitions or the petitions with the most signatures contributing to the effects observed), or users following links shared via email and social media, which would point to a specific petition that the contact was supporting (who will avoid these information cues). These alternative effects might be tested using an experimental approach in future research. In March 2012, the Cabinet Office introduced a change to the e-petitions site which also shaped the information environment of prospective petitioners, by introducing a 'Trending e-petitions' with the highest signatures on the front page. The fact that our data was captured both before and after this change provides us with a 'natural experiment' whereby we can test the effect of this change, which will be the next step in this programme of research.

Conclusions

We have found that in online mobilizations, growth tends not to occur, meaning that most mobilizations that are initiated, fail. But where it does, it proceeds in rapid bursts followed by periods of stasis. Such a finding suggests that online mobilizations of the kind covered here could play a role in the more general process of punctuated equilibria in policy-making. For example Jones and Baumgartner (2005) found a high correlation between public concern on an issue and Congressional attention. Our findings here could be even more interesting. In the theory of punctuated equilibrium the media plays a key role in terms of 'lurching' from one issue to another and having a complex feedback relationship with public opinion. But the sort of mobilizations we are looking at here are bubbling up relatively independently of the media, gaining media attention only when they obtain significantly high levels of support—the petition on road-pricing that successfully played a role in obtaining policy change, for example, received a great deal

of media attention once that it reached one million signatories. Research that develops our understanding of the mechanics of this turbulence will be important for scholars and policy-makers alike as collective action continues to move into online settings. If online collective action is characterised by punctuations, then it looks as if such activity could inject a further dose of instability into political systems. If mobilizations follow a pattern of very low levels of attention punctuated by occasional 'spurts' which grow rapidly into full scale mobilizations that merge with other elements of the political system to push policy change on to the agenda and the institutional landscape, then we can expect to see increasing turbulence in contemporary politics, adding to the 'instability' that Baumgartner and Jones (1993; 2005) and their co-investigators have modelled so extensively in previous research.

This paper has, we hope, demonstrated the potential for 'big data' approaches in political science research. The data we report here was automatically and non-obtrusively generated to provide a dataset of real-time transactional data of a kind that has rarely been available to political science researchers before. One of the aims of the research programme of which this analysis forms a part is to develop the methods we have used to both harvest and analyse the data, which require skills and expertise and conceptual approaches that span academic disciplines; of the authors of this paper, one is a physicist, one a computer scientist and one a political scientist. As big data is used more extensively in this kind of research, the ability to work across disciplines in this way will become increasingly important.

Our future empirical work will also explore co-ordination with media coverage and mentions of the petitions on social networking sites (such as Facebook, YouTube, Google Search and Twitter); we are now systematically gathering data on any mention of the petitions for which we have captured data across all these platforms. Much work on online collective action is based on research into a single platform, whereas any online activity tends to involve several. By looking carefully at the timing with which an issue gains attention in different parts of the political system, including the activist activities investigated here, we might get closer to establishing some sort of sequencing of attention. In addition, one way of getting around the lack of causal inference in research of this kind is to carry out experiments, as in Margetts et al (2011, 2012, 2013). Future work will also use an experimental approach to analyse the effect of different information environments surrounding petition websites on petition growth. Designers of web sites that involve civic engagement, such as e-petition sites, must decide what social

information about existing levels of participation to include, for example the numbers of people who have already signed and the timings of when they did so. Other design decisions include whether participants are anonymous (as on the Cabinet Office site) or whether their names are made visible (as on the earlier No. 10 Downing St site and the German e-petitions platform), and whether input from other social media platforms is incorporated into the petition site. Research of this kind, using both 'big data' and experimental approaches, can inform such design decisions in ways that maximise citizens' input to policy debates.

References

Ackland, R. And Gibson, R. (2006) *Hyperlinks and horizontal political communication on the WWW: The untold story of parties online*. The Australian National University.

Aral, S. and Walker, D. (2011) 'Creating Social Contagion Through Viral Product Design: A Randomized Trial of Peer Influence in Networks, *Management Science*.

Baumgartner, F. and Jones, B. (1993) *Agendas and Instability in American Politics*. University of Chicago Press pp 1-298.

Bennett, L. and Segerberg, A. (2011) 'Digital Media and the Personalization of Collective Action: Social Technology and the Organization of Protests against the Global Economic Crisis.' *Information, Communication & Society* (2011) 14: 770-799.

Bennett, L., and Segerberg, A. (2012) 'The Logic of Connective Action', *Information, Communication & Society*, 15 (5): 739-768.

Bimber, B. (2003) *Information and American democracy: Technology in the evolution of political power*. New York: Cambridge University Press.

Bimber, B, Flanagin, A, Stohl, C (2006) 'Reconceptualizing Collective Action in the Contemporary Media Environment', *Communication Theory*, Vol. 15 Iss. 4: 365-388.

Bøg, M., Harmgart, H., Huck, S. and Jeffers, A. M. (2012), Fundraising on the Internet. Kyklos, 65: 18–30. doi: 10.1111/j.1467-6435.2011.00525.x

Borge, R. and Cardenal, A. (2011) 'Surfing the Net: A pathway to participation for the politically uninterested?', *Policy and Internet*, vol 3, issue 1.

Centola, D. (2010) 'The Spread of Behaviour in an Online Social Network Experiment' *Science* 329, 1194.

Croson, Rachel and Jennifer. Shang (2009*).* 'Field Experiments in Charitable

Contribution: The Impact of Social Influence on the Voluntary Provision of Public Goods', *Economic Journal. 119 (540): 1422–1439.*

Escher, T. 2011. *Writetothem.com: Analysis of users and usage for UK Citizens Online Democracy*, London: My Society.

Etling, B. Kelly, J. Faris, R. and Palfrey, J. (2009) *Mapping the Arabic Blogosphere: Politics, Culture and Dissent*, Berkman Center Research Publication NO. 2009-06.

Frey, Bruno S. *and* Stephan, Meier *(*2004*).* Social Comparisons and Pro-social Behavior: Testing Conditional Cooperation in a Field Experiment*, American Economic Review.* 94*:* 1717–1722*.*

Goel, S. Watts, D. and Goldstein, D. (2012) 'The Structure of Online Diffusion Networks', *Proceedings of the 13th ACM Conference on Electronic Commerce*, pp. 623-638.

Gonzalez Bailon, S. Borge-Holthoefer, J. Rivero, A. and Moreno, Y. (2011) 'The Dynamics of Protest Recruitment through an Online Network', *Scientific Reports*, December.

John, P., & Margetts, H. (2003). Policy punctuations in the UK: Fluctuations and equilibria in central government expenditure since 1951. *Public Administration*, 81(3), 411-432. Blackwell Publishing Ltd. doi:10.1111/1467-9299.00354

Jones, B. D., & Baumgartner, F. R. (2004). Representation and Agenda Setting. *Policy Studies Journal*, 32(1), 1-24. Blackwell Publishing Ltd. doi:10.1111/j.0190-292X.2004.00050.x

Jones, B. and Baumgartner, F. (2005) *The Politics of Attention: How Government Prioritizes Problems*, Chicago: University of Chicago Press.

Jungherr, A. Jurgens, A. (2012) 'The Political Click: Political Participation through E-Petitions in Germany', *Policy and Internet*: 2 (4). Pp. 131-165.

Liben-Nowell, D., & Kleinberg, J. (2008). Tracing information flow on a global scale using Internet chain-letter data. *Proceedings of the National Academy of Sciences*, 105(12), 4633. National Acad Sciences.

Lindner, R. and Riehm, U. (2011) 'Broadening Participation Through E-Petitions? An Empirical Study of Petitions to the Germany Parliament', *Policy and Internet*: 3(1), pp. 63-85.

Lupia, A., & Sin, G. (2003). Which Public Goods are Endangered?: How Evolving Communication Technologies Affect The Logic of Collective Action. *Public Choice*, 117(3), 315-331. Springer Netherlands. Retrieved from

http://dx.doi.org/10.1023/B:PUCH.0000003735.07840.c7

Margetts, H., John, P., Reissfelder, S., and Hale, S. 'Applying Social Influence to Collective Action: Heterogeneous Personality Effects' (July 19, 2011, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1892805)

Margetts, H. John, P., Escher, T. and Reissfelder, S. (2011) 'Social Information and Political Participation on the Internet: an Experiment', *European Political Science Review.*

Margetts, H., John, P., Escher, T., & Reissfelder, S. (2009) Experiments for Web Science: Examining the Effect of the Internet on Collective Action, *Proceedings of WebSci2009 Conference*, Athens, 18-20 March.

Mayer-Schoenberger, V. and Cukier, K. (2013) *Big Data: A Revolution That Will Transform How We Live, Work and Think*. John Murray.

Mosca, L., & Santucci, D. (2009). Petitioning Online. The Role of E-Petitions in Web Campaigning. *Political Campaigning on the Web*, 121. transcript Verlag.

Schelling, T. C. (2006). *Micromotives and macrobehavior*. WW Norton & Company.

Smith, G. (2009). *Democratic innovations: Designing institutions for citizen participation*. Cambridge, England: Cambridge University Press

Whyte, A., Renton, A., & Macintosh, A. (2005). e-Petitioning in Kingston and Bristol Evaluation of e-Petitioning in the Local e-Democracy National Project.

Wright, S. (2012) Assessing (e-) Democratic Innovations: "Democratic Goods" and Downing Street E-Petitions", *Journal of Information Technology and Politics*, 9: 453-470.

Wu, F. and Huberman, B. (2007) 'Novelty and collective attention', *Proceedings of the National Academy of Sciences of the USA,* 104 (45) 17599-17601.