Networks of Communities and Communities of Networks in Online Government

Paul Henman[1], Rob Ackland[2]and Tim Graham[1]
[1]The University Of Queensland, Brisbane, Australia
[2]Australian National University, Canberra, Australia
p.henman@uq.edu.au
robert.ackland@anu.edu.au
t.graham1@uq.edu.au

Abstract: Over the past decade the World Wide Web has become a core platform for the electronic operation of government. Yet the shape and nature of government presence on the Web and the online community in which it resides remains poorly understood and under-theorised. This paper analyses large-scale web crawling data that map the hyperlink network structure between government websites and the broader Web ecology in the UK. In particular, it reports the 'communities' of websites within a hyperlink network of over 19,000 websites and over 135,000 hyperlinks derived from 75 key UK government seed sites at national, regional (i.e. Scotland and Wales) and local government levels. Website communities were derived by utilising Infomap, a state-of-the-art community detection algorithm that operates on the principle that flows of information in complex networks reveals community structure. Identifying and analysing online communities in which government websites reside provides insights in how hyperlink communities are arranged, that is, their emergent organizing principal and the importance of government in these online communities. It is hypothesized that online 'communities' can occur around different policy topics (such as health, education or policing), or along institutional or jurisdictional boundaries (such as England, Scotland and Wales). Using this novel approach this paper demonstrates that communities emerge on both axes, and that social media and government portals are some of the most significant communities based on information flows. This research provides foundational knowledge about the role of government websites in the World Wide Web, the emergent online associations, and the changing dynamic of state information in the twenty-first century. It points to strategies for developing government Web presence in networks that matter.
Keywords: social network analysis, community detection, hyperlink networks, Infomap, web social science, UK

1 Introduction: Government in (online) networks

The rapid emergence and evolution of the internet since the 1990s has generated a plethora of research on governments' use of digital information and communication technologies, especially the internet. A key development in e-Government research has involved a conceptualization and analysis of websites, including the now classical notion of stages of e-Government (Lee, 2010; Siau and Long, 2005). This conceptualization has underpinned international and intra-national rankings of e-Government by the United Nations and others (Lörincz, et al, 2010; UN 2014; West, 2005), and research seeking to identify the drivers for e-Government advancement (Lee, et al, 2012; West, 2005).

Yet despite this substantial body of research on government on the web, very little is known about the nature of government websites vis-à-vis other websites. In short, much of the analysis of government websites treats the sites themselves as individual items for investigation. This perspective is somewhat strange given that a primary and arguably definitive characteristic of websites is that they are simply an ordered network of webpages within a wider network of webpages (and websites) joined by hyperlinks. Focusing e-Government analysis on the network in which government websites are located, instead of the individual sites, raises a number of new possibilities. What is the structure of the network in which government websites are located in, and how do governments seek to structure their network of websites? What role do government websites take in the wider web? How easy is it to find government websites for specific purposes, such as public sector information, laws and policies, government decisions, accessing public services, and initiating complaint and appeal processes, and are these purposes also available through other government websites? How does the online structure of government relate to offline government as defined constitutionally, organizationally or topic-wise? There is already a small body of research examining different elements of these concerns (see for example, Escher et al, 2006; Whalen, 2011; Henman et al, 2014).

This paper seeks to advance such knowledge by examining the different online networks or 'communities' in which government websites are located and their role in those networks or communities. We use the word

Reference this paper as Henman P, Ackland R and Graham T. "Networks of Communities and Communities of Networks in Online Government" *Electronic Journal of e-Government Volume 12 Issue 2 2014, (pp117-130)*, available online at www.ejeg.com

'communities' carefully to refer to groups of websites highly connected by hyperlinks. As such, the relationships are electronically defined, and not by social, legal or organizational relationships. To be sure, such web communities are not communities of people connected online, such as Facebook friends or members of an online discussion group. Rather, they are purely defined by hyperlinks within webpages within websites. As such the construction of a hyperlink to another website will occur for a range of purposes. In relation to government websites, we can imagine that hyperlinks within government sites will reflect the needs of the government agency that develops and controls that site. Hyperlinks to government sites could be expected to reflect a range of rationales based on the purpose of the website and its 'owner'. Accordingly, it is hypothesized that online 'communities' may coalesce around different policy topics (such as health, education or policing), along institutional or jurisdictional boundaries (such as national, provincial, local) or even along different functional purposes (such as web content management, financial services).

Analysing e-Government in terms of online communities, or networks of websites, is timely and engages with a wide contemporary and diverse discourse about networks. Since Castells' influential work, *The Rise of the Network Society* (1996), it is commonplace to conceptualise our contemporary social, economic and organisational world in terms of networks. The network metaphor emphasizes the flattening of power and dynamic spaces of flows, and contrasts with modern constructs of hierarchy, rationality and rigidity. Network frame of mind has been adopted as a normative approach for reshaping organisations for greater creativity, responsiveness and democracy. This theme is also taken up in the governance literature whereby changes in public sector management have been characterized as a shift from governance by hierarchy, to markets, to networks (Jessop 1998; Kjaer, 2004; Rhodes 1997). The network metaphor has also been utilized as a normative frame to advance collaborative governance of public services. In understanding public policy processes and policy change, the notion of networks has been deployed in highlighting the social networks of key human and organizational agents to mobilise or stifle change (Marsh and Rhodes, 1992; Weible et al, 2009). Each of these different network discourses imply different organising rationales for the linkages: along economic functional lines; along service delivery lines; and on political/policy lines. Underpinning some of these network discourses is a recognition of the re-organising capacities made possible by digital ICTs, and the internet in particular.

Yet, the network as a conceptual or analytical concept is not without critique (see for example, Dowding 2001; Kendall, 2004). Key criticisms have been that in its emphasis on relationships, which are often viewed in horizontal or flattening ways, the network imaginary often overlooks or underplays power. Similarly, 'network' is often used as metaphor with limited capacity to distinguish between different networks in terms of constituents and the patterning of relationships within it. Recent developments in social network analysis inspired by the internet and informed by mathematics, have provided new avenues through which to delineate and characterize networks and communities within networks (eg. Ackland, 2013; Easley and Kleinberg, 2010). This paper draws on some of these new methods to provide foundational knowledge about the networks in which government websites are located and the role they play within them.

This paper addresses three main research inquiries based on a hyperlink network surrounding a set of key British national, regional and local government websites. Firstly, we seek to understand the different web communities identified in a network neighbourhood of UK government websites, including what types of websites are dominant in each community. Secondly, we examine the role government websites play in each community and the types of government website that are dominant in them (categorised by tier of government), focusing specifically on the most important communities in the network. Thirdly, we consider the web communities among government websites, that is, exclusive of commercial and other non-government sites.

The remainder of the paper is structured as follows. In Section 2, we summarise the various algorithmic approaches to community detection that are available, and explain the nature and rationale for using the Infomap community detection algorithm. The third section explains the research methodology, including the underlying dataset used. Section 4 then reports the research findings in terms of the structure of the online network vis-à-vis online communities, the nature of those communities and the presence and role of government websites in individual communities and in the community network. The fifth section repeats this analysis for the sub-network made up entirely of government websites to identify the dominant government sites within the government online community. The paper concludes with a reflection on what this might mean for e-Government research and government website development.

2 Networks and 'community detection'

A vast array of natural and digital systems can be represented topologically as networks and examined to identify underlying, and often unforeseen, dynamics and structures. Hyperlinks are the 'fabric of the web' (Helmond, 2013: 3). In this study we analyse a hyperlink network of UK government websites and the websites connecting to them (that is, their neighbours), hereafter referred to as the 'UK e-Gov network'. In the UK e-Gov network, nodes represent websites and edges represent hyperlinks from one website to another. Hyperlinks function similarly to citations, meaning that if a website has an 'outlink' to another website this implies that the other website contains something of value (i.e. information). One can also view hyperlinks as a form of information flow or information navigation. Thus, a website with many 'inlinks' from other websites is generally regarded as authoritative or important. Web users are more likely to 'surf' to such websites because there are multiple pathways provided by the hyperlink structure of the network as a whole (i.e. many other websites linking to a single website). The size and complexity of large networks, such as the UK e-Gov network, makes it difficult to examine and analyse effectively. In order to address this problem, 'community detection' was utilized to identify the natural clusterings, or communities, within the large network as defined by high hyperlink connections. In general, the aim of community detection is to reduce complex networks into modules that "simplify and highlight the underlying structures and the relationships that they depict" (Rosvall and Bergstrom, 2008: 1118).

There are a number of different approaches to 'community detection' in complex networks, including modularity maximization (Lancichinetti and Fortunato, 2012), Edge-Betweenness (Girvan and Newman, 2001), Fast-Greedy (Clauset et al, 2004), Multi-level (Blondel et al, 2008), Walktrap (Pons and Latapy, 2005), Infomap (Rosvall, Axelsson and Bergstrom, 2009) and BNEM (Hazef et al, 2014). However, only three of these approaches support the analysis of graphs that are both directed and weighted: Edge-Betweenness; BNEM; and Infomap. In the UK e-Gov network, edges are directed denoted by a hyperlink from the source website to the receiver website. These edges are weighted according to how many hyperlinks there are from a website to another website. Whereas many community detection algorithms tend to function over the 'underlying' graph (i.e. disregarding the directionality of edges), analyzing the UK e-Gov network requires an approach that takes into account both the direction and weight of edges. Infomap, a cutting-edge approach to community detection in networks, was utilized not only because it supports directed and weighted networks, but it also scales well.

2.1 The Infomap approach: Community detection in hyperlink networks

Infomap is an information theoretic approach to detecting community structure in complex networks. Community detection decomposes networks into 'modules', or communities, according to regularities in network structure. Infomap approaches this task by undertaking the equivalent of random walks along edges. To illustrate, imagine a 'walker' who is placed onto a node in the network and proceeds to walk randomly from node to node. The random walker can only walk to neighbouring nodes via a directed edge; the directedness of edges defines the possible paths of movement because each edge provides a one-way path, or hyperlink, between neighbouring nodes, which may or may not be reciprocated. At any given node the random walker is randomly assigned a path from all possible outwards edges from the current node, where each edge is probabilistically weighted according to its edge weight. In short, there is a higher probability of walking to a node if the edge has a higher weight (e.g. if a website contains 100 out-links to another website rather than just 1 or 2 out-links). These random walks, or flows, are then described by a code that seeks to maximally compress the description of network flow that reveals groups of nodes, or routes, "among which information flows quickly and easily" compared with the rest of the network (Rosvall and Bergstrom, 2008: 1118).The resulting groups of nodes are the modules or communities within which the random walker spends a relatively long time before exiting. In short, the central premise of Infomap is that flows of information via hyperlinks in complex networks reveal structure within a larger network. This enables us to "focus on how the structure of the extant network constrains the dynamics that can occur on that network" (Rosvall and Bergstrom, 2009: 14).

In relation to the hyperlink networks under examination in this paper, Infomap produces communities, or sub-networks, of websites within a larger hyperlink network between websites. These communities of websites are connected because of greater hyperlink connectivity to each other relative to other websites and communities of websites. To be sure, these hyperlink networks are equivalent to road infrastructure, rather than traffic on those roads. They define possible routes that a web user may take, but not the amount of web users taking those particular routes. As such, the communities resulting from this analysis cannot be

interpreted as a group of inter-connected online individuals or levels of information or communication exchange between organisations. At the same time, the resulting hyperlink network communities are analysed to see if they may reflect or relate to other logics, for example organisational connections, cognate policy or service areas, similar geographical locations, shared organisational types.

Algorithmically, the Infomap algorithm outputs information that describes the best two-level partition (shortest description length) of the attempts to partition the network. Concretely, this returns a list of communities, where each community is a list of nodes, and in our case, websites. The communities are ordered according to the total amount of flow between the websites in a community as a percentage of total flow between websites in the entire network. Therefore, while communities with large numbers of members (or websites) may be expected to generate a larger proportion of total flow, a highly dense community with fewer websites could account for a greater proportion of total network flow. Infomap also lists nodes (i.e. websites) within each community from highest to lowest flow volume within that community, on the supposition that those nodes with greater flow volume are more important in defining the nature of the network or can be viewed as having a central tendency in the information flows within that community. The Infomap algorithm also returns a list of edges between the communities weighted by flow volume between each community. In our case, this network of communities is a network of groups of websites.

3 Method

3.1 Generating the network data

The sheer size of government online makes is difficult to examine the entire network of government on the Web and the network in which it is located. Moreover, the structure of government on the Web would be expected to be different for different government jurisdictions (national, regional, local, supra-national) and policy areas, including between countries. In order to analyse the online hyperlink network of government it is necessary to generate such a network. There are a range of techniques, sampling approaches and data gathering methodologies involved in generating hyperlink networks. However, most start with a set of seed sites from which to start the webcrawling and hyperlink generation process.

Given the rationale for this study is to understand the online network of government websites and their relationship with other government and non-government sites, the research presented in this paper generated a hyperlink network from a large, purposively sampled list of key UK government websites from three tiers of government (national, regional and local), multiple policy domains, and central, line and ancillary government agencies and websites. Our approach extends previous work that focused only on national government agency websites (e.g. Escher et al, 2006; Whalen, 2011) to also include regional government (specifically, Scotland and Wales) and local government, as well as the Greater London Authority. Within England, Scotland and Wales, the websites of two rural and two urban local government authorities were included. In total, a selection of websites from key government agencies covering a diverse range of policy and public service areas were identified as reflective of British government on the web. Specifically, the seed sites include central government (the Prime Minister's, Treasury and Parliamentary sites), six policy and public service domains (foreign affairs/defence; health; community services made up of social security and housing; education; environment; law/policing), and government portals. In total, 75 British government websites were used as seed sites with which to generate a larger network of websites (see Appendix).

Using these 75 websites, a hyperlink network was created in September 2012 by identifying hyperlinks out of each webpage in each seed website and all hyperlinks coming into each webpage of each seed site. Hyperlinks between webpages of non-seed websites were also collected. Hyperlink network data were collected and assembled via an iterative process that broadly occurred in two stages. Stage One involved webcrawling each seed site using the VOSON system, a web-based software systems incorporating web mining and data visualisation tools (http://voson.anu.edu.au/). This web crawl collected both 'outbound' links (web pages that the seed sites link out to) and 'inbound' links (web pages that link in to the seed sites). Inbound links were collected via Blekko API. Further, the internal links for each webpage of each seed site were crawled up to a maximum of 1500 webpages in one website. Stage Two involved finding the outbound links for all webpages that were discovered during Stage One (these webpages are referred to as the 'first ring'). Hyperlinks between webpages in the first ring were mapped during this process, resulting in a network of webpages as nodes. In

short, the resulting network was a network of individual webpages with hyperlinks as directed edges between these webpages.

After all these webpages and their hyperlinks were identified, a process of grouping webpages into websites occurred. Such 'pagegroups' were created from the webpages in the network to reflect natural online groupings constituted as websites. Webpages were grouped according to domain names. This process was undertaken to enable network analysis between websites, based on a domain name, rather than between individual webpages. Thus, each node in the network is a collection or grouping of webpages (generally a single domain name). The resulting network, where each node represents a separate website/domain name, contains over 19,000 websites and over 135,000 hyperlinks. To be sure, while all (or most) webpages were collected for each seed site, the pagegroups or websites contained in ringset one comprise of only a selection of webpages that arose from the webcrawling. This network is defined as a "1.5-degree egonet" (Ackland, 2013, p. 50) representing hyperlinks between seed sites and hyperlinks between sites directly connected as neighbours to seed sites (i.e. sites in the 'first ring'), but not the neighbours' neighbours.

3.2 Generating 'communities' using Infomap

The Infomap algorithm was then applied to this large hyperlink network. The Infomap algorithm allows various parameters to be specified which may affect the output. We used two parameters: (1) specifying that the network is *directed*; and (2) specifying N, the number of attempts to partition the network. At each iteration, the Infomap algorithm progressively seeks to partition and sub-partition the network as the random walker walks through the network. The partitioning process is undertaken in order to minimise description length within communities. The latter parameter can significantly affect the results. The algorithm was run for N of 10, 100, 1000 and 20,000. While there was some variation in the results with different values of N, only minor differences were detected with values of N 100 and greater. We report results for N=100.

4 Results: The community structure of the UK e-Gov hyperlink network

4.1 Enumerating and naming communities

The Infomap algorithm generated 996 unique communities from the UK e-Gov network, and resulted in 21,150 links between communities. The median number of websites per community is 2. The maximum number of websites in a single community is 1495 (Community 1– '*Infosphere*'). There are 665 communities that contain three or fewer websites.

A first observation in the communities derived from the analysis is that a relatively small number of communities summarise or make up much of the entire network. Although 996 communities were generated, a very small percentage of communities accounted for most of the overall flow in the full network. Indeed, the distribution of flow between communities in the UK e-Gov network follows a power law distribution. This is often referred to as a 'scale free network' (Barabasi and Bonabeau, 2003), which has been observed more broadly on the Web, whereby a relatively small number of popular websites account for a majority of the world's internet traffic. Similarly for the UK e-Gov community network, the top 5% (50 nodes out of the total 996) account for approximately 90% of all flow. Similarly, the top 100 (10%) nodes account for almost 97% of all flow. Table 1 provides a list of the Top 25 communities, presents the communities in descending order by 'flow' volume between communities (loosely interpreted as 'importance'), and also lists the number of websites (or nodes) in each community.

A second component in the analysis was to understand the nature of each community, in terms of shared characteristics within each community. Identifying how to meaningfully label each community is both critically important as a first analysis step. Naming helps inform us about 'what is going on' in each community. As Rosvall and Bergstrom write that "useful maps assign unique names to important structures" (2008, p. 1118). In practice we observed that a considerable number of communities contain a heterogeneous assortment of websites that do not readily suggest a clear category or label. While Infomap automatically labels communities after the website that has the highest flow volume in each unique community, this can be a useful heuristic, but it also can be misleading.

Table 1: Top 25 communities (descending order by flow volume)

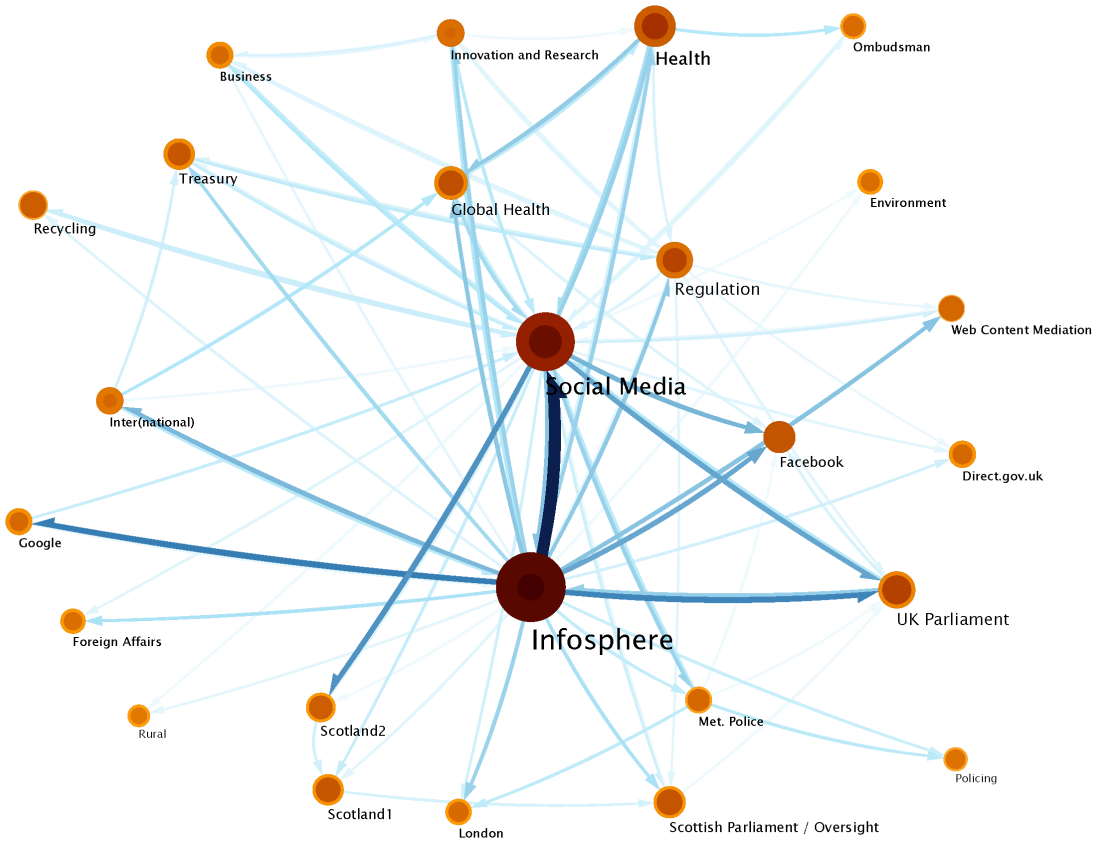
Rank	Community	Aggregated flow volume (%)	Size (number of websites)
1	Infosphere [Blogosphere/News/Wikipedia]	9.0496	1495
2	Social Media	6.7264	271
3	Health	3.7796	417
4	UK Parliament	3.1768	169
5	Regulation [Law/Taxation]	3.0971	170
6	Global Health	2.7099	161
7	Scottish Parliament / Oversight	2.5398	381
8	Facebook	2.5296	2
9	Treasury	2.4551	161
10	Scotland1	2.4467	527
11	Scotland2	2.2467	642
12	Recycling	2.228	631
13	Innovation and research	2.0326	153
14	Inter(national)	2.0013	285
15	Direct.gov.uk	1.9833	323
16	Web Content Mediation	1.9778	38
17	Business	1.948	342
18	Google	1.9465	54
19	Met. Police	1.9449	506
20	London	1.8888	647
21	Ombudsman	1.8316	146
22	Environment	1.8153	390
23	Foreign Affairs	1.7736	562
24	Policing	1.6695	432
25	Rural	1.5257	194

In labelling the communities, we took into account the nature of the websites focusing particularly on the top dozen or so that contributed the most information flow within the community. Consideration was given to the policy and service domain of the websites, the tier of government the sites are associated with, the generic top level domains (country code, and website type – commercial (.co/.com), government (.gov/.go), organisational (.org), etc.), as well as critical awareness of the role and dynamics of key websites and how they relate to the rest of the community. In some cases, there were clearly identifiable themes. For example, Community #3 *Health*, has the British Department of Health (www.dh.gov.uk) as its top site and a large constellation of other health websites, both government and other (.org and .co), with five of the top 10 being British National Health Service (NHS) websites (.nhs.uk). In other cases, there appeared to be two distinct and not obviously related themes within the community, and so the community was doubled-labelled. For example, Community #7 *Scottish Parliament/Oversight* contains a high proportion of websites specifically relating to Scottish Parliament with www.scottish.parliament.uk the principal website with about 9 times the information flow of the second website which is the British Ministry of Justice (www.justice.gov.uk). Other Scottish parliament related sites include www.scotlandoffice.gov.uk and www.audit-scotland.gov.uk. The latter site overlaps with the Oversight theme in this community that includes ombudsman, information commission and public standards websites. In some cases communities were very heterogeneous having no real theme, but when a few websites accounted for the overwhelming majority of the information flow within the community, these few websites determined the label. An example of this is Community #15, *Direct.gov.uk,* which contained the government portal as its principal member, and 50 times the information flow of the next most significant member, and the next nine sites are quite heterogeneous. Therefore, naming communities requires a fair degree of qualitative decision making.

Based on this naming convention, Figure 1 is a network visualisation of the resulting communities in the UK e-Gov network. It displays the Top 25 communities (as per Table 1) and the amount of 'flow' between them. The size of a community (the 'circles') is proportional to the average time a random walker spends moving between websites within that community. The width of a link (the lines between communities) is proportional to the per step probability that a random walker moves between the communities.

In Figure 1 we observe that the Community #1 *Infosphere* (which includes websites such as www.data.gov.uk, www.wordpress.com, www.wikipedia.com, www.guardian.co.uk and www.reddit.com) has a clear central role in the network of communities. Not only does it account for the highest information flow within the UK e-Gov network (i.e. 9%), but it also has the highest outward information flow from it to other communities, including *Social Media* (#2), *UK Parliament* (#4), *Facebook* (#8), *Inter(national)* (#14) and *Google* (#18). *Social Media* (includes websites such as www.twitter.com, www.youtube.com and www.linkedin.com) similarly has an authoritative and central role in the overall community structure accounting for (7%) of overall flows. However, unlike *Infosphere*, the *Social Media* community has a lot of inflows as well as outflows. The relatively symmetric links from *Social Media* to the surrounding communities suggests that *Social Media* plays a critical role in routing information and traffic to and from separate communities (e.g. *Health, Business*, *Recycling*, and *Met. Police*).

Figure 1: Visualizing the community structure of UK e-Gov network (Top 25 communities)



4.2 The makeup of UK e-Gov network communities

Apart from using naming as a heuristic in analysing the communities in the UK e-Gov network, quantitative measures of website characteristics were also calculated for each community. This analysis identifies which communities are largely delimited by country or jurisdictional boundaries or organisational type. This further enriches our understanding of the makeup of communities. Table 2 provides a breakdown of the Top 25 communities by *Country Code Top Level Domain* (i.e. .uk, .au, .de) and *Generic Top Level Domain* (e.g. .gov, .com, .org). It is important to note that in the table, 'Gov' sites includes all sites with a clear government code (namely .gov, .go, .govt, .gv and .gouv). In addition, it also includes all UK sites with .police, .parliament and .nhs as their generic top level domain as they are clearly government sites. The top level domain .sch, which is for schools, is included in 'Other' as not all schools are government run.

We observe that the communities are typically dominated by UK websites or unknown country code, the latter of which are websites with no country code, that is, ending in .com, .org and so on. This result makes sense given the seed sites are British. The main exception is *Inter(national)* (#14), which has 46 percent of the community with sites country codes other than the UK. This reflects the community's name. Perhaps

unsurprisingly, *Social media* and *Infosphere* communities (#1, #2) have over half of their websites being unknown, and this reflects the high level of commercial sites in these communities, as indicated in the second last column. Whereas communities with a majority of British websites tend to reflect a government function (i.e. #4, #5, #9, #15, #17), a policy or public services domain (i.e. #3, #13, #17, #22, #24, #25) or a geographical domain (i.e. #10, #11).

Table 2: Percentage of websites in Top 25 communities by country code and web domain

	Community	Country Code TLD (%)			Generic TLD (%)				
		UK	Other	Unknown (e.g. .com, .org)	Gov	Org	Net	Com	Other
1	Infosphere	16	18	66	2	20	7	54	18
2	Social Media	42	6	51	3	20	2	67	8
3	Health	70	4	26	29	36	2	27	6
4	UK Parliament	49	4	47	14	40	5	33	9
5	Regulation	53	5	42	16	17	1	63	4
6	Global Health	30	20	50	8	32	2	40	19
7	Scottish Parliament / Oversight	41	10	49	3	29	3	49	15
8	Facebook	0	0	100	0	50	0	50	0
9	Treasury	49	7	44	11	35	1	45	8
10	Scotland1	58	4	38	11	42	2	38	7
11	Scotland2	53	4	43	2	35	3	51	9
12	Recycling	45	7	48	2	20	2	68	8
13	Innovation and research	69	3	28	13	31	2	29	25
14	Inter(national)	8	46	46	13	15	2	35	35
15	Direct.gov.uk	64	5	31	12	39	2	39	7
16	Web Content Mediation	45	8	47	8	21	8	61	3
17	Business	63	1	36	6	10	2	80	3
18	Google	43	7	50	7	22	2	57	11
19	Met. Police	28	15	57	4	20	5	58	14
20	London	35	9	56	7	23	4	55	11
21	Ombudsman	77	3	21	20	40	2	32	5
22	Environment	59	4	37	10	26	2	58	5
23	Foreign Affairs	39	16	45	6	16	2	65	11
24	Policing	74	2	25	15	13	2	67	3
25	Rural	51	8	41	7	31	3	49	10

In terms of Generic Top Level Domain, government websites (.gov, etc.) do not constitute a significant proportion of any community of the top 25 communities. Rather, websites in the commercial (.com/.co) or organizational (.org) domain make up the large bulk of each community, which is not surprising given global statistics of domain registrations. Communities in which government websites show a solid presence are *Health* (29%), *Ombudsman* (20%), *Regulation* (16%) and *UK Parliament* (14%). What we can conclude is that government websites, even in very government orientated functions such as *UK Parliament, Treasury* and *Ombudsman*, are a small proportion of the communities. However, this should not be interpreted that they are small players in such communities, as they often account for significant information flows within each community.

Table 3: UK Government websites in Top 25 communities by Tier of Government or Jurisdiction

	Community	National	Greater London	N. Ireland	Scotland	Wales	Local	Police	Schools
1	Infosphere	5	0	0	2	0	14	0	1
2	Social Media	6	1	0	2	0	0	0	0
3	Health	122	3	0	2	0	1	0	0
4	UK Parliament	7	1	0	0	0	1	0	0
5	Regulation	11	2	0	1	1	0	0	0
6	Global Health	6	2	0	0	0	0	0	0
7	Scottish Parliament / Oversight	7	0	1	10	0	1	0	2
8	Facebook	0	0	0	0	0	0	0	0
9	Treasury	13	0	0	0	0	0	1	0
10	Scotland1	7	0	0	31	0	28	0	0
11	Scotland2	0	0	0	25	0	3	1	2
12	Recycling	2	0	0	1	2	0	0	0
13	Innovation and research	18	2	0	0	0	1	0	0
14	Inter(national)	0	1	0	0	0	0	0	0
15	Direct.gov.uk	7	1	0	0	3	10	0	0
16	Web Content Mediation	3	0	0	1	0	0	0	0
17	Business	9	0	0	0	1	1	0	0
18	Google	3	1	0	0	0	0	0	0
19	Met. Police	1	1	0	0	0	0	3	1
20	London	1	11	0	0	0	30	0	0
21	Ombudsman	23	0	3	1	2	1	0	0
22	Environment	11	0	2	0	2	5	0	0
23	Foreign Affairs	3	0	0	1	0	0	0	0
24	Policing	3	1	0	1	1	2	57	0
25	Rural	6	0	0	0	0	2	0	0

In seeking to understand further the role of government websites in the top 25 communities in the UK e-Gov network, the websites were coded according to British government jurisdiction (i.e. national, Scotland, Wales, Northern Ireland, Greater London Authority and local government). Sites with .nhs were coded as national, but those with .police and .sch (school) were separately classified. Table 3 provides the results of this breakdown. Some communities exhibit fairly strict jurisdictional boundaries. For example, *Treasury, Health, UK Parliament* and *Foreign Affairs* largely include only national UK government websites, and *Scotland2* consists primarily of Scottish government websites, while *Policing* is dominated by sites with .police as their top level domain name. In contrast, some communities challenge jurisdictional boundaries. *Scottish Parliament/Oversight* contained sizeable number of websites from both national and Scotland jurisdictions reflecting its dual naming, *Scotland1* included strong numbers from Scotland and local government (which makes logical sense) and *London* has an understandably strong showing of Greater London Authority and local government. The *Environment* community includes government sites from national, Northern Ireland, Wales and local government, which is reflective of the ubiquity of environment policy at all government levels. Overall, there appears to be a logic to these results, but not one that could be predicted, only interpreted.

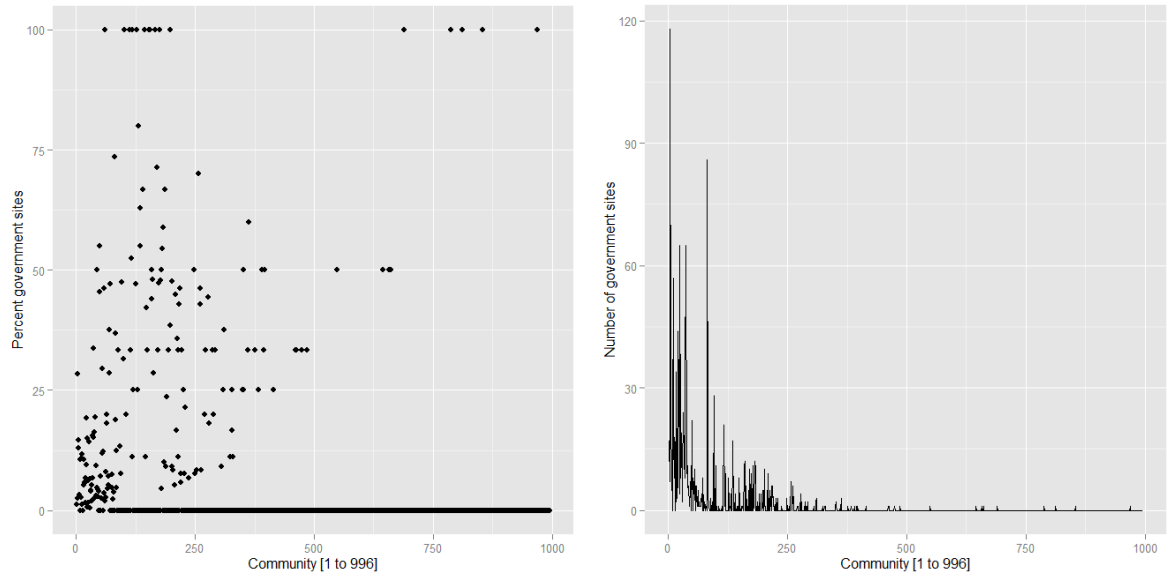


Figure 2: percentage (2a, LHS) and number (2b, RHS) of UK government websites in each community

The foregoing analysis provides insights into the role and makeup of government websites in the top 25 communities that accounts for two-thirds of total information flow in the network. In order to understand whether or not government websites overall are playing an important role in the full network, it is worth considering how many of the government websites in the entire network are in the more significant communities or spread in some other fashion. There are a total of 1513 UK government websites (or about 8%) within the entire UK e-Gov network. Figure 2(a) illustrates the percentage of UK government websites in each community as a percentage of the total number of websites in each community, while Figure 2(b) plots the actual number of UK government websites in each community. Recall that communities are ordered in terms of importance in ascending order (i.e. Community 1 is most important). Across both figures we observe a strong pattern of UK government websites towards the left-hand side of the graph. This indicates that, UK government websites are largely located in the most important communities in the UK e-Gov Web ecology.

5 Conclusion

This paper presented foundational knowledge about the community structure in which UK government websites are located. Using 75 British government websites from national, regional and local levels across a range of policy domains a wider network of over 19,000 websites and over 135,000 edges were generated. This network consists of all the (inwards and outwards) hyperlink neighbours to the seedsites, and the hyperlinks between these neighbours. Using Infomap, almost 1000 communities were identified in the network to reflect the information flows defined by the weighted and directed edges within the entire network. Yet only a small percentage of these communities are of any significance; out of 996 communities, the top 5% account for over 90% of all flow throughout the network, and 10% account for 97% of flow. The question for e-government researchers and administrators is what role government websites play within this network of communities. Do they provide a strong central or organising role? Are they present within the top communities? On what basis do communities cluster, for example around specific policy domains, public service types, jurisdictions or otherwise?

In addressing these questions, the communities resulting from Infomap algorithm on the UK e-Gov hyperlink network were analysed in various ways. Overall, government websites play a key role in the resulting network of communities. At the same time, there are key communities where government websites are clearly not significant, such as *Social Media* (#2), *Facebook* (#8), *Web Content Mediation* (#16) and *Google* (#18).

The types of websites within each community are generally quite diverse, including a mix of Generic Top Level Domains (e.g. .com, .gov, .org) and Country Code Top Level Domains, yet generally a dominant theme or two could be discerned among the websites in each community (see Table 1). Of the British government websites within each community, there are diverse combinations of tiers of government and policy domains. However, a small number of communities exhibit fairly concise jurisdictional or policy boundaries (e.g. *Health, Policing*

and *Ombudsman*). Furthermore, government websites tend to be included in the most important communities of the large network, based on flow volume. This applies to both British government (i.e. gov.uk, .parliament.uk, sch.uk, police.uk and nhs.uk) and non-British government websites.

These findings indicate that the Infomap algorithm is highly suited to this kind of research and provides a useful conceptual tool to examine the 'social life' of large-scale Web hyperlink networks, including analysing the role of government websites within the web. Our initial expectations about the structure of government on the Web are partly supported by the results, which reveal that to some extent communities do form around different policy and public service topics or institutional and jurisdictional boundaries. However, there are at least two caveats to this hypothesis. Firstly, such communities are generally quite diverse in terms of their constituent websites (i.e. they often transgress institutional boundaries of policy domain and jurisdiction). Secondly, whilst important communities formed around government seed sites, we also found that these communities competed equally for visibility and dominance with non-government communities. Furthermore, communities included a mix of both government (.gov) and non-government websites (.org, .net, .com), challenging the notion that government on the Web is structured similarly to traditional 'offline' arrangements. Indeed, government sites never make up a large component of websites in a community, though this is perhaps not surprising, given the makeup of the World Wide Web, and the population of offline organisations.

Given that only about half of the top 25 communities are not strongly government orientated, it is crucial that e-Government schemes continue to forge strategic links and relationships across the commercial and non-profit Web. In particular, social media and blogging communities are extremely important players in the overall Web ecology. Yet is it understandable that government websites are not located within such communities even if governments make extensive use of such online functions. Governments will benefit from building in-roads and out-roads within, and between, these communities, and working to increase their significance within each community. Moreover, the success of e-Government Web strategies can be quantified by measuring the extent to which government websites are positioned in terms of community structure as a whole, and within each community. That is, whether government websites are strategically networked at the centre of information flows on the Web, which we have shown can be examined and visualized in terms of communities.

Given the strong movement to the creation of government web portals in recent years it is useful to consider the role these play within the network of communities and communities of networks (von Lucke, 2007). Indeed, the British government has had two iterations of web portal development. At the time of the data collection for this study, the website www.direct.gov.uk was the British government principal portal, but was progressively replaced from October 2012. There are also several other British government portals that were included in the seed sites, namely www.businesslink.gov.uk and www.data.gov.uk, www.scotland.gov.uk and the Welsh government and Greater London Authority websites are largely a single website/portal (www.wales.gov.uk and www.london.gov.uk). Significantly, all the national portals are within the top 25 communities – direct.gov.uk in #15, businesslink.gov.uk in #17 and data.gov.uk in #1 – and each are the number one website in each community. This is reassuring to government and not surprising given that portals are designed as large websites with wide coverage of information. As such they are expected to attract hyperlink attention from external websites, but also may well act as hubs pointing to external websites. Similarly, the Scottish and Greater London Authority portals are the principal website in communities #10 and #20 respectively. Perhaps surprisingly, the Welsh government portal appears as second website in the #32 community, while Northern Ireland's portal's lowly status as fourth in community #36 could result from it not being used as a seed site.

To sum up, we offer two points for consideration. Firstly, governments are continually expanding and evolving their Web presence in order to improve and optimize e-Government projects. Community detection of government hyperlink networks provides a useful and novel tool for strategic e-Government analysis and management toolbox. Analysing community structure provides a unique window into the changing nature of state information in the 21st century; governments can examine the Web-networked relationships between various arms of the state and the broader Web ecology and act on this knowledge to repair, create and strengthen linkages. Finally, as governments have moved to centralise existing government websites into 'one-stop shop' portals, this affects the community structure of government on the Web as anticipated by their designers. Given that our web crawling data were collected in late 2012, future studies might examine what kinds of community structures have since emerged from the ongoing

evolution and portalisation of government websites and the implications for e-Government.

Acknowledgements

We gratefully acknowledge the support of the Australian Research Council Discovery Grant (DP110100446), and the technical computer expertise contributed by Lin Chen in the first half of this project. We also gratefully acknowledge the insightful and technical assistance from Martin Rosvall (InfoMap creator) in successfully utilising InfoMap.

References

6, P. (2004) *E-governance: styles of political judgment in the information age polity*, Basingstoke: Palgrave.

Ackland, R. (2011) 'WWW Hyperlink Networks', in Hansen, D. L., Schneiderman, B. and Smith, M. (ed.) *Analyzing social media networks with NodeXL: Insights from a connected world.* Amsterdam: M. Kaufmann.

Ackland, R. (2013) *Web social science: Concepts, data and tools for social scientists in the digital age*, London: Sage publications.

Barabasi, A. and Bonabeau, E. (2003) *Scale-free networks,* New York: Sci American Inc.

Blondel, V. D., Guillaume, J., Lambiotte, R. and Lefebvre, E. (2008) 'Fast unfolding of communities in large networks', *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10): P10008.

Clauset, A., Newman, M. E. J. and Moore, C. (2004) 'Finding community structure in very large networks', *Physical Review .E, Statistical, Nonlinear, and Soft Matter Physics*, 70(6 Pt 2): 066111.

Dowding, K. (2001) 'There must be an end to confusion: policy networks, intellectual fatigue, and the need for political science methods courses in British universities', *Political Studies*, vol. 43, pp. 136-158.

Easley, D. and Kleinberg, J. (2010) *Networks, crowds, and markets: Reasoning about a highly connected world*, New York: Cambridge University Press.

Edler, D. and Rosvall, M. (2013) *The MapEquation software package*, [Online], Available: http://www.mapequation.org.

Escher, T. and Margetts, H. (2006) 'Governing from the Centre? Comparing the Nodality of Digital Governments', [Online], Available: http://ssrn.com/abstract=1755762

Girvan, M. and Newman, M. (2002) 'Community structure in social and biological networks', *Proceedings of the National Academy of Sciences of the United States of America*, 99(12), pp. 7821-7826.

Helmond, A. (2013) 'The Algorithmization of the Hyperlink', *Computational Culture,* 3, [Online], Available: http://computationalculture.net/article/the-algorithmization-of-the-hyperlink

Henman, P. (2013) 'Government and the Internet: Evolving Technologies, Enduring Research Themes', In Dutton, W. H. (ed.) *The Oxford Handbook of Internet Studies,* Oxford: Oxford University Press.

Henman, P., Ackland, R. and Graham, T. (2014) 'Social Policy on the Web: The Online Institutional Structure of Social Policy Domains in the UK', *Proceedings of the 14th European Conference on e-Government (ECEG '14)*, 12-13 June, Brasov Romania.

Hudson, J. and Lowe, S. (2004) *Understanding the policy process: Analysing welfare policy and practice*, Bristol, U.K.: Policy Press.

Jessop, B. (1998) 'The rise of governance and the risks of failure: the case of economic development', *International Social Science Journal*, *50*(155), pp. 29-45.

Kendall, G. P. (2004) 'Global networks, international networks, actor networks', In Larner, W., and Walters, W. (ed.). *Global governmentality: governing international spaces*, Routledge, pp. 59-75.

Kjaer, A. M. (2004. *Governance: key concepts,* Cambridge: Polity.

Lambiotte, R. and Rosvall, M. (2012) 'Ranking and clustering of nodes in networks with smart teleportation', *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*, 85(5 Pt 2), 056107.

Lancichinetti, A. and Fortunato, S. (2011) 'Limits of modularity maximization in community detection', *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*, 84(6 Pt 2), 066122.

Lee, J. (2010) '10 year retrospect on stage models of e-Government: A qualitative meta-synthesis', *Government Information Quarterly*, *27*(3), pp. 220-230, [Online], Available: http://unpan3.un.org/egovkb/en-us/Reports/UN-E-Government-Survey-2014.

Lee, C. P., Chang, K. and Berry, F. S. (2011) 'Testing the Development and Diffusion of E-Government and E-Democracy: A Global Perspective',*Public Administration Review*, *71*(3), pp. 444-454.

Lörincz, B., Tinholt, D., Linden, N. v. d., Colclough, G., Cave, J., Schindler, R., et al. (2010) *Digitizing Public Services in Europe: Putting ambition into action*, EC, Directorate General for Information Society and Media, [Online], Available: https://ec.europa.eu/digital-agenda/sites/digital-agenda/files/egov_report.pdf.

Marsh, D. and Rhodes, R.A.W. (1992) *Policy networks in British government*, Oxford: Clarendon Press.

Pons, P. and Latapy, M. (2005) *Computing communities in large networks using random walks*, Berlin: Springer-Verlag Berlin, pp. 284-293.

Rhodes, R.A.W. (1997) *Understanding governance: Policy networks, governance, reflexivity and accountability*, Buckingham: Open University Press.

Rosvall, M. and Bergstrom, C. T. (2008) 'Maps of random walks on complex networks reveal community structure', *Proceedings of the National Academy of Sciences of the United States of America*, 105(4), 1118-1123.

Rosvall, M., Axelsson, D. and Bergstrom, C. T. (2009) 'The map equation', *The European Physical Journal Special Topics*, 178(1), pp. 13-23.
Siau, K. and Long, Y. (2005) 'Synthesizing e-government stage models - a meta-synthesis based on meta-ethnography approach', *Industrial Management + Data Systems,* 105/3-4: pp. 443-458.
von Lucke, J. (2007) 'Portals for the Public Sector', in Anttiroiko, A.-V. and Mälkiä, M. (ed.) *Encyclopedia of Digital Government*, Hershey, P.A.: Idea Group Reference, pp. 1328–1333.
Weible, C. M., Sabatier, P. A. and McQueen, K. (2009) 'Themes and variations: Taking stock of the advocacy coalition framework', *Policy Studies Journal*, *37*(1), pp. 121-140.
West, D. M. (2005) *Digital government*. Princeton, N.J.: Princeton University Press.
Whalen, R. (2011) *The Structure of Federal eGovernment: Using hyperlinks to analyse the .gov domain*, Paper 8, [Online], Available: http://opensiuc.lib.siu.edu/cgi/viewcontent.cgi?article=1009&context=pnconfs_2011.

Appendix

	UK National/England	Scotland	Wales	Greater London Authority
Central	Government Portal www.direct.gov.uk	Scottish Portal www.scotland.gov.uk	Wales Portal wales.gov.uk/?skip=1&lang=en	London Portal london.gov.uk
	Parliament www.parliament.uk	Parliament www.scottish.parliament.uk	Welsh Assembly assemblywales.org	London Assembly london.gov.uk/assembly
	Business Portal www.businesslink.gov.uk	Scottish Cabinet scotland.gov.uk/About/14944/Scottish-Cabinet	First Minister Cabinet wales.gov.uk/about/firstminister	Mayor london.gov.uk/mayor
	Prime Minister www.number10.gov.uk	Office of the Chief Economic Adviser scotland.gov.uk/Topics/Economy/EconDept/OCEAEcon	Cabinet Members wales.gov.uk/about/cabinet/cabinetm	Budget and Performance Committee london.gov.uk/moderngov/mgCommitteeDetails.aspx?ID=129
	Data Portal data.gov.uk		Directorate for Strategic Planning, Finance and Performance wales.gov.uk/about/civilservice/directorates/spfp	Budget (Mayoral Responsibility) london.gov.uk/who-runs-london/greater-london-authority/budget-and-strategic-plan
Legal	Ombudsman www.ombudsman.org.uk			
	Treasury www.hm-treasury.gov.uk			
	Her Majesty's Revenue and Customs www.hmrc.gov.uk			
Foreign Affairs/ Defence	Foreign and Commonwealth Office www.fco.gov.uk			
	Ministry of Defence www.gov.uk/government/organisations/ministry-of-defence			
Health	Department of Health dh.gov.uk	Health and Social Care Directorate scotland.gov.uk/About/Directorates	Directorate of Health and Social Services wales.gov.uk/about/civilservice/directorates/hsscdirectorate	Health and Public Service Committee london.gov.uk/moderngov/mgCommitteeDetails.aspx?ID=148
	Health Protection Agency (Public Health) hpa.org.uk	NHSScotland show.scot.nhs.uk	Wales Health wales.gov.uk/topics/health	Health Priority london.gov.uk/priorities/health
	NHS			

	nhs.uk			
	NHS LiveWell nhs.uk/LiveWell			
	Sector Regulator of Health Services in England monitor-nhsft.gov.uk			
	NHS Institute for Innovation and Improvement institute.nhs.uk			
Community Services	Department for Work and Pensions dwp.gov.uk	Governance and Communities Directorate scotland.gov.uk/About/Directorates	Community Services wales.gov.uk/topics/socialjustice	Equalities Priority london.gov.uk/priorities/equalities
	Jobcentre Plus direct.gov.uk/en/Employment/Jobseekers/ProgrammesandServices/index.htm	Scottish Commission for Children and Young People sccyp.org.uk	Children and Young People wales.gov.uk/topics/childrenyoungpeople	
	Child Maintenance Enforcement Commission childmaintenance.org		Older People/Social Services wales.gov.uk/topics/olderpeople	
Education	Department for Education education.gov.uk	Learning and Justice Directorate scotland.gov.uk/About/Directorates	Directorate for Education and Skills wales.gov.uk/about/civilservice/directorates/educationandskills	
		Education Scotland educationscotland.gov.uk	Department for Education and Skills wales.gov.uk/topics/educationandskills	
Environment	Department for Environment, Food and Rural Affairs defra.gov.uk	Enterprise and Environment Directorate scotland.gov.uk/About/Directorates	Directorate for Sustainable Futures wales.gov.uk/about/civilservice/directorates/sustainablefutures	Environment Priority london.gov.uk/priorities/environment
	Environment Agency environment-agency.gov.uk	Scottish Environmental Protection Agency www.sepa.org.uk	Environment and Countryside wales.gov.uk/topics/environmentcountryside	Environment Committee london.gov.uk/moderngov/mgCommitteeDetails.aspx?ID=143
	WRAP (Waste and Resources Action Programme) wrap.org.uk		Department of Environment and Sustainable Development wales.gov.uk/topics/sustainabledevelopment	
Law / Policing	Attorney General attorneygeneral.gov.uk	Scottish Police scottish.police.uk	Attorney General attorneygeneral.gov.uk	Crime and Community Safety Priority london.gov.uk/priorities/crime-community-safety
	Home Office homeoffice.gov.uk/police	Office of the Advocate General advocategeneral.gov.uk	Police wales.gov.uk/topics/housingandcommunity/safety/police	
	UK Police police.uk			
Housing		Housing scotland.gov.uk/Topics/Built-Environment/Housing	Department for Housing, Regeneration and Heritage wales.gov.uk/topics/hou	Planning and Housing Committee london.gov.uk/moderngov/mgCommitteeDetails.

			singandcommunity	aspx?ID=158
		Scottish Housing Regulator scottishhousingregulator.gov.uk		Housing Priority london.gov.uk/priorities/housing
Local Govt (urban)	Birmingham www.birmingham.gov.uk	Edinburgh www.edinburgh.gov.uk	Newport City newport.gov.uk	
	Leeds www.leeds.gov.uk	Glasgow glasgow.gov.uk	City and County of Swansea swansea.gov.uk	
Local Govt (rural)	Ribble Valley ribblevalley.gov.uk	Western Isles cne-siar.gov.uk	Isle of Anglesey anglesey.gov.uk	
	West Somerset www.westsomersetonline.gov.uk	Scottish Borders scotborders.gov.uk	Vale of Glamorgan valeofglamorgan.gov.uk	