Prepared for delivery at the 2006 Annual Meeting of the American Political Science Association, August 31 - September 4, 2006

Governing from the Centre?

Comparing the Nodality of Digital Governments

Tobias Escher,
Research Fellow, Oxford Internet Institute, University of Oxford
& School of Public Policy, UCL

Helen Margetts
Professor, Oxford Internet Institute, University of Oxford
Honorary Senior Research fellow, School of Public Policy, UCL

Vaclav Petricek,
Research Fellow, Department of Computer Science, UCL
& YAHOO

Ingemar Cox,
Professor of Telecommunications,
Departments of Computer Science and Electronic & Electrical Engineering, UCL
Director, UCL's Adastral Park Campus

Governing from the Centre?
Comparing the Nodality of Digital Governments

Abstract

What difference does e-government make to the capacity of governments to interact with citizens? How does it affect government's place in social and informational networks - the 'nodality' of contemporary government? What is the structure of 'government on the web' and how do citizens experience government on-line? This paper uses methods from computer science (particularly webmetrics) and political science (a 'tools of government' approach) to go further than previous work in developing a methodology to quantitatively analyse the structure of government on the web, building on Petricek et al (2006). It applies structural metrics (via webcrawling) and user metrics (via user experiments) to the web sites of comparable ministries concerned with foreign affairs in three countries (Australia, the US and the UK). The results are used to assess the on-line presence of the three foreign offices along five dimensions: visibility, accessibility, extroversion, navigability and competitiveness. These dimensions might be developed further as indicators for use by both researchers (to assess e-government initiatives) and by governments (to improve the effectiveness and efficiency of their on-line presence). Governments which are successful in developing their web sites in this way are likely to have greater visibility to citizens, businesses and other governments, strengthening nodality as a policy tool.

Governing from the Centre?
Comparing the Nodality of Digital Governments

The governments of most advanced industrial nations have an on-line presence, amounting to thousands of web sites and many millions of web pages in even medium-sized states. But what is the structure of this electronic arm of government and how accessible is it to citizens wishing to interact with government? How does a government's on-line presence affect that government's policy-making capacity?

We might expect that rising use of the internet and related technologies would increase one key tool of government policy: 'nodality', the extent to which government is at the centre of social and informational networks (Hood, 1983; Hood and Margetts, 2006). Yet web sites vary in the extent to which they are accessible, visible and connected to other parts of the internet. Differences across and within sectors (governmental, voluntary and commercial) could mean that governmental organisations find themselves competing for nodality in the on-line world.

This paper uses methods from computer science as well as political science and social network analysis to go further than previous work in developing a methodology to quantitatively analyse the nodality of government on the web, building on Petricek et al (2006). First, it discusses the concept of nodality and how it might be broken down into key dimensions such as visibility and navigability. Second, it applies two key categories of metrics to the web sites of comparable ministries concerned with foreign affairs in three countries (the US, the UK and Australia): structural metrics and user metrics. For structural metrics, we look at the extent to which a site is successful in disseminating information to the outside world, via incoming links from other sites (analogous to 'authority' in web metric terms) and the extent to which a site offers its users links to alternative sources (analogous to 'hubness') of information and expertise. We also use metrics from social network analysis to make a comparative assessment of the navigability of each site, providing objective indicators of how easily we predict information can be found once inside the site. For user metrics, we report on a number of user experiments carried out by the authors to establish how users actually 'experience' the foreign office sites when searching for government-related information.

This paper looks at information seeking – rather than service transactions (which generally involve authority, treasure (money) or organisation, the other 'tools' of government policy). While previous work has in general paid more attention to the narrower focus of transactional services, we

believe that the delivery of information should be regarded as a service in its own right, as well as being an important part of any other transaction – just as it is in the commercial sector. Certainly, by 2006 citizens are far more likely to interact with government on-line when looking for information (24 per cent of UK citizens according to Dutton et al 2005) than when conducting transactions such as payment, form-filling or registration or making contact with representatives (9 per cent). Governments which are successful in maximising the extent to which citizens can and do easily find government-related information on-line are likely to have greater visibility to citizens, businesses and other governments, strengthening nodality as a policy tool. Conversely, governments which are less successful may find themselves becoming increasingly 'invisible' as they compete for attention with organisations from private, public and voluntary sectors pursuing more successful strategies, thereby suffering a net loss of nodality and ultimately, policy-making capacity.

Nodality as a 'tool' of government

Various attempts have been made to identify the 'tools' of government policy. The one used in this paper was developed by Christopher Hood (1983) and has been recently revised for the digital age by Hood and Margetts (2006). Hood identified the 'tools' of government as follows:

- **Nodality:** the property of being at the centre of social and informational networks
- **Authority:** the possession of legal or official power to demand, forbid, guarantee, adjudicate
- **Treasure:** the possession of a stock of money or exchangeable goods
- **Organisation:** the possession of a stock of people with whatever skills they may have (soldiers, workers, bureaucrats), land, buildings, materials, computers and equipment, somehow arranged

These four tools (and combinations thereof) form the basis of the toolkit available to governments seeking policy solutions to societal problems. The tool we are concerned with here is 'nodality'. Nodality 'equips government with a strategic position from which to dispense information' (Hood, 1983:12); the greater a government's nodality, the more likely that it can use the dissemination of information alone to change societal behaviour, rather than resorting to the use of the more costly tools of authority, treasure or organisation. Even where these other tools are employed, effective use of nodality will maximize efficiency. Government's presence on-line offers the first opportunity to measure nodality. While proxy measures for the other tools have long been used and argued over, nodality was, until the advent of the internet and the worldwide web, virtually impossible to quantify.

With widespread and rising use of the internet by citizens and governments, we might expect government to become more nodal as the Internet and associated technologies become more embedded into all aspects of social and political life. However, if private sector organizations and non-governmental organizations are more successful at using the World Wide Web to increase their nodality, it may be that government will suffer a net loss of nodality in the virtual realm. We can hypothesize that a 'healthy' government domain, if we can establish appropriate characteristics to define such a thing, will help government to become more nodal. If a domain has more incoming links, for example, it is likely to be more visible to search engines and more easily found by citizens searching for government-related information.

Most advanced industrial nations have put considerable political support and financial resources behind the development of e-government. By 2005, the UK for example has a '.gov' domain of around 8 to 23 million pages (depending on which search engine estimates one tends to believe, MSN or Google respectively) and was spending £14.5 billion a year on information technology in the pursuit of the Prime Minister's commitment to have all government services electronically available by the end of 2005. In spite of these resources (greater than 1 per cent of GDP in most industrialized nations is spent on government information technology), e-government tends to lag behind e-commerce. In the UK, recent survey evidence (Dutton et al, 2005) suggests that while 85 per cent of Internet users claim to have looked for or bought goods and services online, and 50 per cent of users to shop online at least once a month, only 39 per cent have had any sort of interaction with government online in the last year. While figures for e-government usage are much higher in some countries, particularly Scandinavia, the generalization that government has been far less touched than commerce by widespread use of the World Wide Web holds true internationally. Governments are under pressure to demonstrate that the massive investments they are making are worthwhile.

So why this mismatch between e-government and e-commerce and why might we hypothesise that government nodality could wane in the on-line world? It is not difficult to find examples where poor visibility of government on-line could lead to a weakening of government's capacity to interact with citizens and even a lack of achievement of policy aims. For example, from 2001 the US Federal Bureau of Investigation (FBI) have advertised on the 'Most Wanted' section of their web site the reward of £25 million offered by the US Department of State for information

leading to the apprehension and conviction of the terrorist Osama Bin Laden[1]. But a Google search for Bin Laden in 2006 does not return the FBI site (or that of any other federal department) in even the top 20 returns (beyond which, research suggests, most users do not access). In contrast, the top 20 listings included various media reports, an entry on Bin Laden's life in the free on-line encyclopaedia (Wikipedia), the full transcript of videotapes sent to the Arab TV station Aljazeera in 2004 and 2001[2], the full text of the terrorist's Fatwa against the American peoples (issued in 1996 and 1998) [3] and a collection of Anti-Osama flash games and movies to either download or play online from a private sector media company.

Likewise, governments face competition from other governments on-line. A striking example of this kind of competition occurred after a massive Tsunami hit south-east Asia on 26th December, 2004. In western countries, concerned citizens automatically turned to their governments for information about missing relatives who had been travelling in affected areas, but many found that other governments were providing more reliable information (Swedish and British governments in particular were criticised for reacting 'way too slow' (*Los Angeles Times*, 31st December 2004). Indeed, it was Internet bulletin boards, operated by voluntary organisations such as the International Red Cross, private sector travel organisations such as Lonely Planet and media organisations that became widely-used sources of information for those seeking news about foreigners in affected countries these organizations operated as a huge information exchange between those on the ground and those seeking news of relatives from afar.

Governments also face competition from a myriad of 'cyber-crime' organisations or individuals, who tend to be sophisticated in maximising their on-line presence. Hackers even maximise their own nodality by advertising their skills on web sites, while operating outside the jurisdiction in which their presence is most obvious. The site http://www.ug100.com/ for example, gives rankings to an astonishing range of underground sites, including those which offer hacking devices, password breakers, credit card generators and a variety of illegal products and services. So for example, if you are seeking information on 'stun guns' on the web in the UK, the first site to appear on a search engine listing will be a site offering to sell you one, rather than any authoritative information. Likewise, a recent scam was run successfully in Ireland on by an organisation offering to apply for a European Health Insurance Card for Irish citizens travelling in the EU, at the cost of

[1] see http://www.fbi.gov/wanted/topten/fugitives/laden.htm

[2] at http://english.aljazeera.net/al

[3] http://www.mideastweb.org

£15, when in fact the service was available on-line for free on the site of the Health and Safety Executive. By obtaining good rankings in the search engine listings, the site was able to 'beat' the official site and many citizens ended up paying for a service they could easily have had for free.

These various types of competition can weaken government's own ability to give out information to citizens. Even where the competition comes from a quasi public organisation, government can find its own capacity weakened. For example, the place where parents seeking information on the comparative performance of schools in the UK ('school league tables') are most likely to visit is the BBC web site, rather than that of the Department for Education and Skills (where the information is also held but in less visible or accessible form). There is no problem with the BBC providing this information – but if citizens are book-marking that part of the BBC site and visiting it regularly, if the DfES did want to disseminate school-related information of another kind, they would find it more difficult to do so, as citizens are not accustomed to visit their site.

Clearly, these examples in themselves do not amount to a definitive threat to government's place 'at the centre of social and informational networks'. But if we were to find systematic evidence that governmental organisations (in specific sectors or countries) are losing visibility on the internet relative to the rest of organisational life, it might also suggest that nodality was waning.

Methods

So how might we measure nodality? This section suggests five dimensions of nodality, each of which would be assessed using a range of 'webmetric' indicators. First, we might assess the 'visibility' of an organization on the web, particularly in terms of the likelihood of it being returned in search engine results. Second we might assess its 'accessiblity'; once we have found it, how easy is it to obtain the information we were looking for. Third, how 'navigable' is it; how easy is it to move around the site easily and quickly and find related content? Fourth, how 'extrovert' is the site, in terms of the number of other sites and sources of information to which users are directed once they are there? Fifth, how competitive is the site, in terms of competing against other sites which provide similar information.

We chose to test our methodology on a single agency example – the foreign office or department of foreign affairs – which may be compared across several countries. The UK Foreign and Commonwealth Office, the US Department of State and the Australian Foreign Office all have roughly comparable roles and responsibilities. For all, nodality could be described as the dominant tool of the organization when interacting with the outside world, rather than authority, treasure or

organisation. We believe that performing a comparative evaluation of their online presence will take us some way towards developing a methodology for comparing government sites in general and even the much larger task of comparing whole government domains. It builds on our previous paper (Petricek et al, 2006) which used a similar methodology to compare the on-line presence of National Audit Offices in five countries (Canada, the Czech Republic, New Zealand, the UK and US) and we will refer to the results of that paper here.

One methodological challenge with this type of research is to define what actually encompasses a website of a foreign office. The foreign offices quite regularly distribute their services across a number of sites, most notably providing a separate domain name for visa & passport services and travel advice. While we have to analyse whether this distribution will actually increase navigability / nodality, we first of all have to make an informed decision on whether to sum up several "web sites" to one FO website or whether to treat them separately. This will always be a qualitative decision. In order to make it as objective as possible we followed two rules:

1. Is the service offered supposed to be an FO responsibility?
2. Is the site operated by the FO?

Table 1 shows the departments and the corresponding websites that were included in our sample.

Table 1: Departments and corresponding websites included in our sample.

Country	Department	Domains
Australia	Department for Foreign Affairs and Trade	http://www.dfat.gov.au http://www.smartraveller.gov.au http://www.passports.gov.au http://www.trademinister.gov.au/ http://www.foreignminister.gov.au/
United Kingdom	Foreign and Commonwealth Office	http://www.fco.gov.uk http://www.ukvisas.gov.uk
United States	State Department	http://www.state.gov http://www.unitedstatesvisas.gov

We excluded embassy websites. Embassies are a huge additional source of data which are operated rather independently from the main site, and are usually indicated by another address. Indeed, only for the US are some embassy websites hosted under the state.gov umbrella. In contrast to the other two countries, the US has at least a common embassy domain (http://www.usembassy.gov).

To quantify our dimensions of nodality we used a number of web metrics, which may be categorized as either structural (and non-obtrusive) or user-based, as follows:

Structural metrics measure properties such as the average distance between two random pages and the interconnectedness of sites. They are readily available to anyone capable of crawling websites and seem to offer the potential for establishing the 'health' of a domain. For example, if a government domain is highly inter-connected, then citizens are much more likely to find information (such as how to make an application for a visa) by traversing the link structure of the site, as the site will be more easily 'navigable'. If a high number of other sites link to the domain, then citizens are more likely to find the site when using external search engines and the site is more 'visible'. If the site has a high number of links pointing out, we might determine it to be more 'extrovert', in the sense that it provides the user with additional information. In addition to basic link structure, mathematical graph theory and social science network analysis have established a range of further measures to describe structural properties of networks which have been applied to web sites (see the following section for references). One such measure is whether there exists a connection between any two randomly chosen documents on that site in a way that allows navigating from one page to another by following hyperlinks. In other words, is there a path between document A and document B – and how long is it? Figure 1 illustrates the concept.

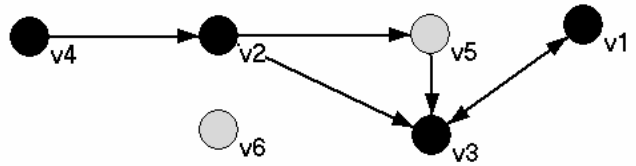

Figure 1: Illustration of simple graph with shortest path between v4 and v1. Note that there exists another path from v4 to v1 via v5 which is not the shortest path and that there is no path back from v1 to v4. There are no paths at all for v6.

Following from these general observations, we identified a number of specific measures that we wanted to compute for each web site, as outlined in Table 2, with the intention of using these to assess the five key dimensions of nodality identified above, all of which might shape a governmental organisation's ability to interact with citizens:

- Visibility – the extent to which the site is easily found using search engines
- Accessibility – the extent to which a search return for the site will take the user 'deep' inside the site, or will deposit them at the top page
- Extroversion – the extent to which the site provides the user with a wider range of information than that held within the site, providing links to other sectors for example

- Navigability – the extent to which the user can easily find relevant content within the site

The last column of the table shows which of these dimensions of nodality each metric might be used to assess.

Table 2: Measures of the structural properties of websites

Name	Description	Interpretation	Indicator of:
Inlink analysis	How many links are there into the site? Which types of domain are linking to the sites – governmental, commercial or voluntary and from which countries?	High numbers of in-links will raise site's visibility on search engine listings – making site more visible to users of external search engines	**Visibility**
	Are the links generic (to home page) or specific (to particular topic on site) – 'deep' links	'Deep links' returned by search engines tend to be more useful – as take users directly to information required rather than depositing them at home page where they must navigate to find what they want – but note deep links more likely to break	**Accessibility**
Outlink analysis	How many links to external sites are there? Which type of domains (governmental, commercial or voluntary) do sites link?	High numbers of external links indicate an 'extroverted' rather than 'introverted' site which leads citizens to a wide range of further information, enriching the interaction	**Extroversion**
Directed average distance	average length of all shortest path	Indicator of how easy it is to navigate the site. Smaller values indicate it is easier to reach other relevant pages	**Navigability**
Diameter	longest of all shortest paths	Worst case measure of navigability	**Navigability**
Strongly connected component (SCC)	percentage of total pages for which there exists a path between any two pages	The bigger, the better as within this component, every page is reachable from any other page	**Navigability**
OUT component	percentage of total pages that can be reached from SCC, but have no path into SCC	Should be smaller as pages in OUT component constitute a dead-end for navigation – although we caution against penalizing sites with large libraries of PDFs	**Navigability**

Name	Description	Interpretation	Indicator of:
Combined-measure: normalized average distance by SCC	a new measure that we propose: the normalized average distance set into relation to the size of the SCC (in %).	While a short path length is good, on its own it does not take into account how much of the site can actually be reached by that average path length. Sites that do not allow every page to be accessed (that is, SCC is smaller than 100%) should be punished.	**Navigability**
Unreachable pairs	percentage of all possible pairs of pages for which no path exists	Should be small - as many pages as possible should be reachable from every page	**Navigability**
Average degree	average number of links (incoming and outgoing) per page	Indicates how heavily interlinked a site is. Similar to density of a network, but independent of size.	**Navigability Visibility**

User-based metrics measure a variety of characteristics of usage of a site, for example 'hits' or page impressions. Acquisition of user-based metrics on a wide scale however is extremely difficult, since the data is only available to the owner of the website, and the data will be time consuming and costly to obtain across multiple sites. Furthermore, co-operation may not be forthcoming: owners of web sites may not wish their data to be used in evaluative studies which highlight poor usage and may not wish their usage data to be compared with other sites. There is likely to be a skew in any data collected this way, as owners of poorly performing sites are the least likely to make their data available. Clearly, there is a huge range of data collected automatically by search engines and internet service providers but privacy concerns make it very difficult to obtain or use such data for academic research. Note for example, the controversy when the US Department of Justice asked Google for a million anonymised records of users of pornographic sites in 2005 or AOL released the logs of all searches done by 500,000 of their users over the course of three months (March- May) in 2006. The data was 'anonymized', meaning that each screen-name was replaced with a unique number, but did allow at least one individual to be identified and was described by many commentators as a 'blatant violation of users' privacy [4]. Even if such data was easily available, it would tell us a great deal about user behaviour, but much less about the performance of specific sites, as it does not contain information about site navigation.

[4] http://sethf.com/infothought/blog/

We have tackled the problem of obtaining user data by setting up and running our own experiments to see how subjects interact with governments online when looking for government-related information. The object of this user study was to compare the performance of the three foreign office websites in different settings. We were interested in whether subjects would find information more easily by using external search engines, internal search capabilities or by clicking through the site only. Furthermore, when subjects were free to choose, would they use the foreign office websites at all or would they access alternative information providers? These considerations resulted in a 3x3 design with three different treatments applied individually to each country, as outlined in Table 3.

Table 3. The design included three treatments compared for each of the countries. The table highlights the rationales for the individual treatments.

<table>
<tr><th></th><th>Treatment 1
open access to whole Internet</th><th>Treatment 2
access limited to site only – searching allowed</th><th>Treatment 3
access limited to site only – no searching</th></tr>
<tr><td>AU</td><td rowspan="3">Will subjects use the sample sites to answer information? What tools do they use in order to locate information? Who else is providing relevant information?</td><td rowspan="3">How do subjects use the sample websites to find information - do they search or do they navigate? While the information is definitely there, will subjects be able to find it?</td><td rowspan="3">Can subjects find the information on the sample websites by simply clicking through the site without searching? What influence has the different structure of the sites for how quickly subjects will locate information?</td></tr>
<tr><td>UK</td></tr>
<tr><td>US</td></tr>
</table>

In order to test how easy or difficult it was for people to find information from foreign office websites we devised 16 questions that asked for particular information provided on all three foreign office sites in our sample. These questions were put before subjects in three treatments as above. The methods used are described in the Methodological Appendix below. The experiments allowed us to assess how users actually experienced the sites and to assess the 'competitiveness' of the sites in terms of competing with other information providers, the fifth indicator of nodality introduced above.

Previous Work

Current research in computer science, political science or communications research tells us little about the nodality or the web structure of e-government. There have been numerous attempts to assess e-government internationally, in the form of rankings of countries carried out or

commissioned by international organisations (such as UNPAN, 2003; Cap Gemini for the European Commission, 2004), private sector consultancies (particularly by Accenture, 2002-2005; Taylor Nelson Sofres, 2003; and Graafland-Essers and Ettedgui, 2003) and academic commentators (West, 2005; La Porte et al, 2001; Demchak et al, 2000). While some are widely cited and eagerly awaited by governments which score well, most rely, ultimately, on subjective judgments. Most make some form of assessment of government websites according to content (eg. West, 2005) and availability of services (eg. Cap Gemini, 2004), while Accenture's widely known annual study is largely a qualitative analysis based on researcher assessments of websites and available e-services and a limited number of short visits to the 22 countries covered. None of these studies have considered the link structure of e-government sites, or systematically collected user metrics, although some use survey evidence (Taylor Nelson Sofres (2003) in particular, while Accenture included a user opinion survey for the first time in 2005) to estimate the extent to which a population as a whole have interacted with their government online.

Similarly, there have been numerous studies within the computer science community to assess and characterize the structure of hyperlinked environments, but few have been applied to government. The idea of a link as an endorsement, inspired by bibliometrics, has been successfully applied to a wide range of problems from ranking algorithms (Brin and Page, 1998; Page et al, 1998; Gibson et al, 1998; Kleinberg, 1999), through focused crawling (Diligenti, 2000) to web page classification and clustering (Gibson et al, 1998; Kumar et al, 1999; Flake et al, 2002). There have also been extensive studies investigating the structure of the Web (Adamic and Huberman, 2001; Broder et al, 2000; Albert et al, 1999), as well as proposals for its generative models (Hindman et al, 2003; Broder et al, 2000; Chakrabarti et al, 1998; Zhou and Mondragon, 2004), all of which noted the scale-free structure of the network. Usually, the study of hyperlink structure has focused on academic networks (Caldas, 2004; Thelwall, 2003). Studies have benefited greatly from the methods developed for social network analysis (see for example Nooy, 2005) and in recent years researchers from various areas have tried to apply these methods to the Internet by interpreting the relation between actors through the hyperlink connections of their websites (Park, 2003).

The application of computer science methods to the study of politics on the web and e-government in particular is not yet very common, although there are some notable exceptions. For example, Hindman et al (2003) studied the communities surrounding political sites and showed that (i) the number of incoming links is highly correlated with the number of actual users and (ii) that online communities are usually dominated by a few sites – winners who take all the attention. Overall

however, applications of computer science, and especially web metrics, to the quantitative evaluation of e-government have not been reported.

Results

The basic properties of the websites in our sample are shown below in table 4. Navigable content means that all doc, pdf and other documents that were not readable by our crawler are filtered out. For all sites, the navigable content forms a significant proportion of the site.

Table 4: Basic properties of foreign office websites

Country	Number of pages		doc/pdf
	Whole site	navigable content	
AU	32,765	30,690	1,667
UK	23,570	20,867	2,439
US	129,246	115,888	13,091

Inlinks

Following the common interpretation of an external in-link as an endorsement, we looked at the number and type of links that each website received. The total number of inlinks, as reported in Table 5 below, used in can be interpreted as an indicator of a site's visibility or authority. Clearly, the number of inlinks is very high for the UK and US. However, the total number of inlinks does not reflect the fact that some countries have a very much smaller population than others. We argue that government websites have a strong relevance for citizen of the respective country but not so much for 'non-citizens' – citizens of other states. While this argument might be less strong for foreign offices (as compared to tax agencies for example), we still believe that the majority of the target audience belongs to the country (by target audience we mean all the people in a country that have access to the Internet, the Internet population of that country). We therefore believe it appropriate to normalize the number of inlinks to a website by the estimated size of the internet population of the associated country (that is, divide by the population in millions).

Table 5: Inlink Analysis for Foreign Office Sites

	Internet Population (millions)	*No. of external in-links*		*% of pages receiving in-links*	*No. of links to top page*	*Links to top page as % of total links*
		Total	*Normalised*			
Australia	14.2	86,195	6,070	11	26,495	31
UK	37.8	859,201	22,730	8	684,797	80
US	203.8	858.066	4,210	25	386,220	45

Note: The Yahoo site explorer (at http://siteexplorer.search.yahoo.com) was used to estimate numbers of inlinks for each site; all other results were obtained via webcrawling

As we might expect, the UK and the US receive far more inlinks than Australia. Once we do normalize, UK emerges as the most 'authoritative' site, with by far the greatest number of inlinks per citizen. Normalisation improves the situation for Australia, which emerges with a higher number of normalized in-links than the US.

However, when we look at what percentage of these links come to the top page, we see a different picture. For Australia only 31 per cent of inlinks coming to the site point to the homepage suggesting that the other 69 per cent penetrate deeper into the site and may well provide more useful content. In contrast 80 per cent of inlinks to the UK site point to the homepage, suggesting that only 20 per cent provide the user with any more focused information. As we will see from our user study below, this characteristic goes against the way users actually used the sites in experimental conditions.

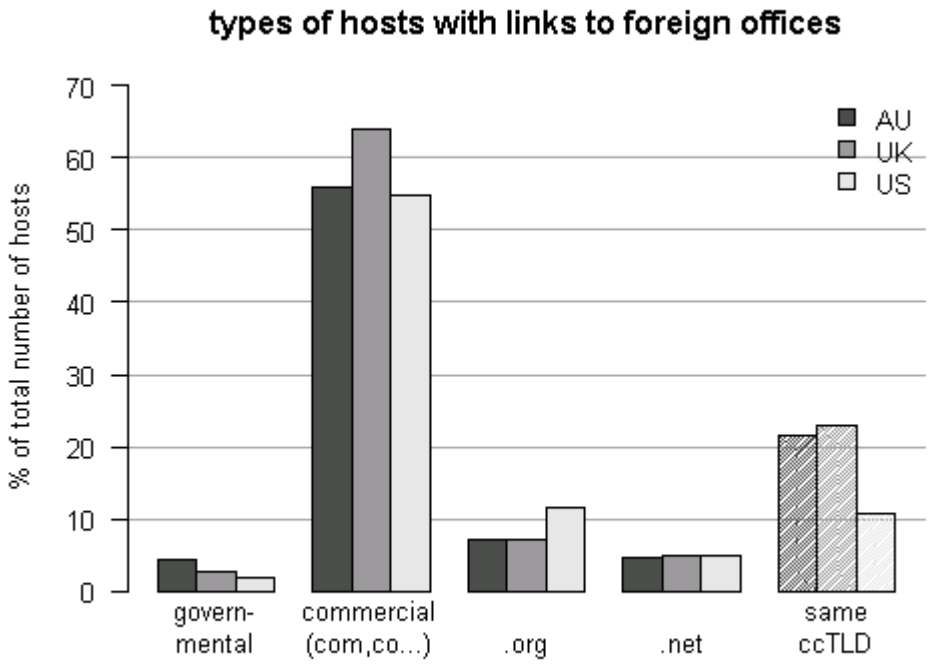

Figure 2: Breakdown of the types of hosts which link to the foreign offices in our sample. Please note that the last column is just an indicator as it is difficult to assign hosts to a certain country if they are under a generic domain name such as .com for example.

Apart from the sheer numbers of inlinks, we carried out a brief analysis of the type of sites linking to the foreign offices, by analyzing the top level domain (the last letters of the domain name, ie. .uk or .com). In an international context, the US site is the most popular one, as the total number of links already demonstrated. In addition, the US receives links from almost every country in the world

(168) and almost 70 foreign governments link to the US State Department, 24 of them are other foreign offices. This is an international and political popularity that cannot be matched by the other two foreign offices. However, there is little difference in the percentage of links received from commercial sites, .org or .net sites. In fact, the majority of websites linking to the foreign offices are commercial (around 60 per cent). It remains to be seen whether this is a feature of government sites in general or whether that is particular to the foreign offices; in some ways, it does not come as a surprise that the majority of links originates from commercial websites, given that these account for the vast majority of websites on the Web, although precise data on the distribution of top level domains is hard to obtain.[5] The figure also shows the percentage of hosts of the respective country that link to the sample website, but this number (generally around 20 per cent) must be interpreted with extreme caution due largely to the fact that .com, .org and .net domains are not easily attributable to a specific country. [6]

Outlinks

We also analysed links from the foreign office sites to other resources, shown in Table 6 below. While online information providers do not have much influence over the people linking to their sites, foreign offices choose the extent to which they disseminate other information to their users and act as a 'hub' in the internet arena. We were interested in the extent to which sites are 'introverted' or 'extroverted' in their own selected links with the outside world.

Table 6: Pages and hosts linked to by foreign office sites.

	Size (No. of pages)	Number of External Links		Number of Different External Pages Linked to		% of Pages on Site with Ext. Link
		Total	Norm-alised	Total	Norm-alised	
AU	32,765	6,017	0.18	2,420	0.074	5.8
UK	23,570	139,357	5.91	60,924	2.585	22.5
US	129,246	120,013	0.93	23,101	0.179	47.0

Note: Where figures are normalized, they have been divided by the number of pages on the site. These normalized values can be read as the average number of pages/hosts linked to per page on the website.

[5] See for example http://www.isc.org/index.pl which reports 171m .net domains and only 69m .com. However, most surveys see the .com clearly on top, for example http://www.domaintools.com/internet-statistics/ with 54m .coms and 7.8m .nets. Similar results from http://www.ipwalk.com/blog/?p=44 (53m .com and 7m .net domains).

[6] This is a problem in particular for the US as its ccTLD (.us) is hardly used and we only counted .us, .mil, .edu and .gov for the US. Many companies in the UK will link to their foreign office but are not attributable to the UK as they use a .com and not a .co.uk domain name. The same applies for .net and .org domains which we cannot attribute to a country.

Although the smallest site after normalisation, the UK seems to be the most extrovert in terms of its links to other pages, both in terms of total numbers and links per page. On average, a page of the UK foreign office has 6 outlinks and links to 2.5 different pages (reflecting the fact that there are a lot of links pointing to the same page). On a per page basis, the UK has by far the highest number of links to other websites (hosts) and this is also reflected in its international scope: the UK has at least one link to 170 different countries, compared to 142 for the US and only 82 for Australia. Interestingly almost every second page on the US site contains outlinks, while only about a quarter of the UK site does so.

Many of the links to 'foreign' websites are to governments and other foreign offices. We can conclude that the UK foreign office has a strong international focus, followed by the US, both in strong contrast to the Australian site. Indeed, Australia is a good example of a very inward looking (or introverted) government site. It has a tiny amount of outlinks in comparison with the other sites, even when normalised, only 6 per cent of all pages contain links outside the site at all and around 40 per cent of them are targeted at other government websites, with the majority targeted at other websites in the AU foreign office portfolio of sites. In contrast, we found that the US and UK are recommending more resources that lie outside the government domain. Both sites link to several thousand other websites of which only around 15-17 per cent are governmental. But in relation to its size the UK is networking much more seriously than the US.

While we could show in our inlink analysis that the majority of sites linking to the foreign offices are commercial, this number drops for the outlinks where the distribution is more evenly balanced. When able to choose – as is the case for outlinks - foreign offices tend to associate more often with governmental sites and less with commercial sites, although of all the websites linked to by foreign offices, about 20-30 per cent are commercial. The UK appears to point to more commercial hosts than the US or Australia: around 35 per cent of hosts, as opposed to just over 20 per cent for the other two.

Network Analysis

We also computed a range of other metrics for our sites, to explore in more depth their internal structure and assess their navigability. The following table reports the measures for our sample websites, as outlined in a previous table. If not mentioned otherwise, the values reported are for navigable content of the site only.

Table 7: Values of metrics for sites in sample. For a definition, see Table 2.

Country	Directed Average Distance (Diameter)	Median Directed Distance	Normalized		Combined-measure: normalized average distance by SCC
			Directed Average Distance (Diameter)	Median Distance	
AU	8.1 (38)	6	1.80 (8.42)	1.33	1.97
UK	4.9 (10)	5	1.12 (2.29)	1.14	1.55
US	6.2 (17)	6	1.21 (3)	1.17	1.45

The following table reports the data for components and average degree both for the whole site and only navigable content. In our sample, the OUT component is equivalent to the percentage of unreachable pairs. This is both due to our crawl, that leaves no other component than SCC and OUT (ie. no IN) but furthermore it indicates that there are almost no links for pages in the OUT component. This table shows that Australia has the largest SCC and the lowest number of unreachable pairs, suggesting at first glance that it might be the easiest site to navigate.

Table 8: Measuring the size of web site component

Country	whole site				navigable content			
	SCC %	OUT %	average degree	Unreach-able pairs %	SCC %	OUT %	Ave-rage degree	Unreach-able pairs %
AU	89	11	54.6	11	95	5	58.1	5
UK	65	35	36.5	35	73	27	40.3	27
US	75	25	38.8	25	84	16	42.6	16

The following figures analyse the distribution of the path length in more depth. In Figure 3 we plot the distribution of the distances (navigable content only) between all pages. This means that for all possible pairs of pages the path length is computed and then plotted according to how many pairs of pages (as percent of the total number of possible pairs) have a path length of less than a certain value. As there are a lot of pairs that do not have a path between them, the graph converges to a value below 100 per cent, usually close to the size of SCC as the pages in the OUT component do not provide many paths to other pages.

In contrast, Figure 4 visualizes the distribution of distances as measured from the top page of a website. So the graph shows how many pages (as percent of the total number of pages) are accessible within a certain number of clicks when starting navigation from the homepage. As our crawl started from the homepage, all pages in our sample are accessible (which is why the graph converges to 100 per cent) but it is interesting to see how many clicks it takes to reach all the sites in

the sample. This value is usually a bit smaller than the diameter of a site. This figure shows that although overall the SCC for AU is 95 per cent, this is only achieved by a very long path length. If we assume that few users will navigate much more than 6 clicks (highlighted in the figure) we see little difference between Australia and the US. By the same token, while the UK has the smallest SCC it is the only one that makes more than 50 per cent (namely 70 per cent) of its content accessible within 6 clicks. This difference between the sites becomes even clearer if we focus on the group of people who would start navigation from the homepage if not using an external search engine. While for the UK, 100 per cent of the content is accessible from the top page within 6 clicks, the same number of clicks would provide only about half of the content for the other two sites. This observation has led us to reconsider our metrics. While a big SCC is of interest, we want a metric that accounts for the number of clicks required to reach pages in this SCC. Several studies suggest (for example, Bernado, 1998) that users are not inclined to follow more than a small number of links, suggesting that the SCC measure may be more useful if it reflects the percentage of pages that are accessible within 6 clicks or so.

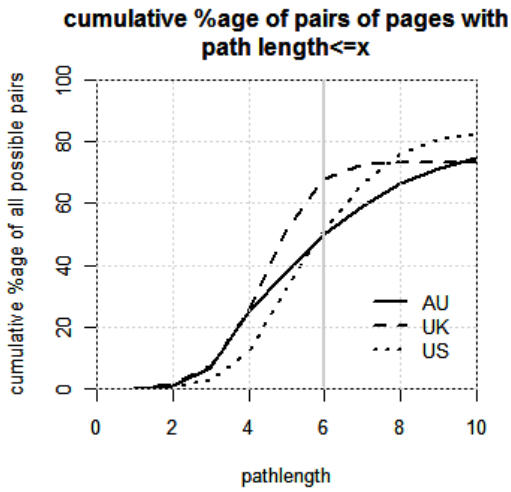

Figure 3: Cumulative sum of number of pairs of pages that have a path between them of less than a certain length. The *x*-axis represents the path length. The *y*-axis represents the percentage of all possible pairs of pages in the website that are connected by a path of length less than x

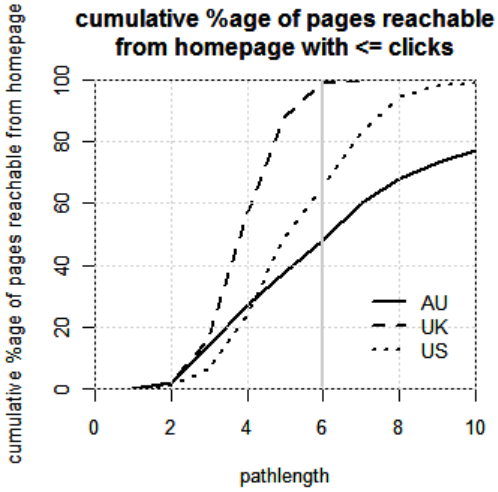

Figure 4: Cumulative percentage of pages that are in a certain path length from the homepage. The x-axis represents the length of the path from the homepage. The y-axis represents the percentage of total number of pages. For the UK, almost all pages are within a distance of 6 clicks reachable from the homepage

Assuming that the majority of users use external search and therefore will not necessarily start at the home page (as was verified in our experimental study, see below) we decided to focus on what percentage of the site is accessible with no more than 6 clicks in general, not just from the top page. This metric yields the following results:

AU: 49.9 per cent,

UK: 68.17 per cent

US: 51.41 per cent (see Figure 3 again).

This metric suggests that the UK will be a more navigable site than Australia and the US, a finding to be tested in our experimental research in the next section.

User Experiments

This section reports on the experiments we carried out to see how subjects actually experienced the web sites in our study. Subjects were provided with a number of information seeking tasks that required specific information provided by the three sites in our sample[7]. Under Treatment 1 subjects could use unlimited search, in Treatment 2 they were restricted to the foreign office site(s) but could use internal search and in Treatment 3 they could not use search at all, but had to navigate. Table 9 gives the basic measures we calculated in order to evaluate how easy or difficult it was for our subjects to find the relevant information, facilitating comparison across the three countries and three treatments.

Table 9: Basic results of experiment compared for all countries and each treatment.

	Unlimited Search Treatment 1			**Internal Search Only Treatment 2**			**No Access to Search Treatment 3**		
	AU	**UK**	**US**	**AU**	**UK**	**US**	**AU**	**UK**	**US**
No. of subjects	15	15	15	18	19	18	11	11	12
Average success rate (correctly answered questions per minute)	0.401 (0.12)	0.438 (0.13)	0.445 (0.09)	0.418 (0.14)	0.411 (0.09)	0.294 (0.06)	0.301 (0.08)	0.458 (0.11)	0.260 (0.09)
%oage of subjects finishing all questions	67%	87%	87%	56%	84%	22%	45%	64%	33%
Average number of answered questions (not skipped)	9.3 (1.2)	9.7 (0.7)	9.8 (0.6)	9.3 (0.9)	9.7 (0.7)	8.5 (1.2)	9.1 (1.0)	9.6 (0.5)	8.2 (1.8)

[7] In our post-experiment questionnaire, subjects identified some questions as not very relevant to foreign office sites and following this feedback it seemed reasonable to exclude a number of low-rated questions in order to focus only on information that met the needs of our subjects, leaving us with 10 questions overall.

	Unlimited Search Treatment 1			Internal Search Only Treatment 2			No Access to Search Treatment 3		
Average path length	6.5 (2.0)	5.7 (1.2)	6.1 (1.4)	7.2 (1.4)	6.0 (1.2)	10.6 (2.3)	7.6 (2.1)	5.3 (1.0)	12.5 (3.4)
Average time to answer a question	147.8 (45.1)	139.3 (38.4)	131.6 (22.0)	148.8 (53.9)	146.3 (29.9)	170.4 (34.0)	169.3 (24.0)	126.0 (25.6)	185.4 (39.4)
Internal search used at least once	80%	60%	53%	72%	95%	94%	*not allowed*		
Average use of internal search (as %age of questions for which it was used)	16% (0.2)	13% (0.1)	19% (0.2)	30% (0.3)	39% (0.3)	35% (0.2)			

Note. Values reported in brackets are standard deviation of the respective mean.

In order to better compare the results and their significance, we include a visualization of the most important measures below in Figure 5. Each plot includes the mean for each country and treatment together with its confidence interval (0.95) so that it can be seen easily whether difference between countries are significant. The general observation is that there are no significant differences in the performance of the three sites when users are free to choose how to find the information. However, the results also show that the subjects in treatment 1 found only about 60 per cent of the information they required to answer on the foreign office sites; the remainder was obtained from other providers. For treatments 2 and 3, in which our subjects were restricted to the sites alone, the link structure and the quality of the internal search engine determined the performance. The overall result is that the UK site performs best, while the US site performs worst. AU occupies the middle ground but loses out as soon as internal search is prohibited.

Our main performance measure is an aggregate of different variables and measures how many questions per minute a subject would answer correctly (success). Figure 5 shows how this measure varied across treatments. Subsequently path length, average time and number of answered questions followed roughly the same pattern. While the internal search was used by almost everybody when limited to the site only, on average it was only used for about a third of the questions. So it seems that subjects tend to try to find information on the site without the help of the search engine, possibly reflecting the quality of internal search engines.

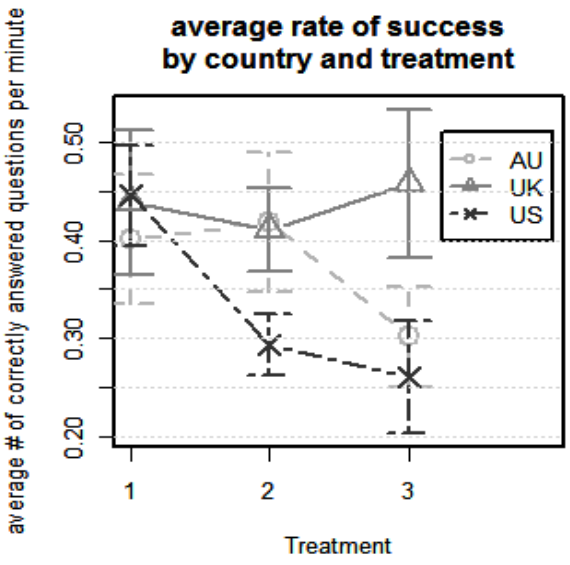

Figure 5a: Average number of correctly answered questions per minute (success) by treatment, plotted for each country with confidence intervals (0.95). While no significant differences can be observed for treatment 1, the UK and AU perform significantly better than the US in treatment 2 and in treatment 3 the UK does significantly better than the rest.

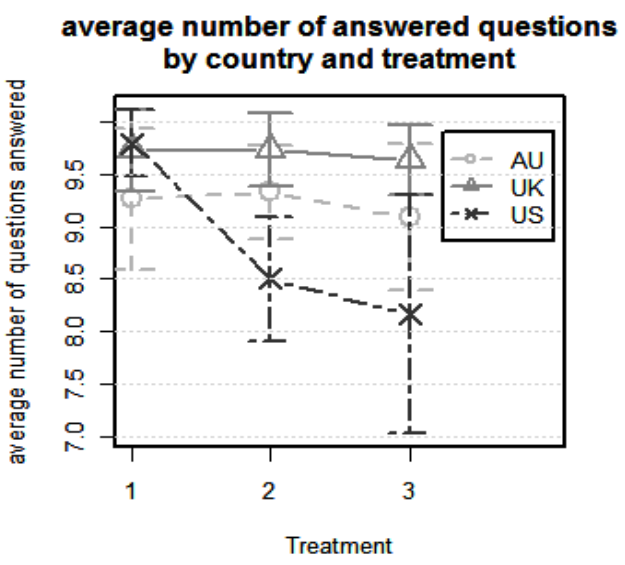

Figure 5b: The average number of answered questions by treatment plotted for each country with confidence intervals (0.95). While there is no significant difference for treatment 1, the UK always outperforms the US in the other two treatments. AU is occupying the middle ground with no significant difference to either of the countries.

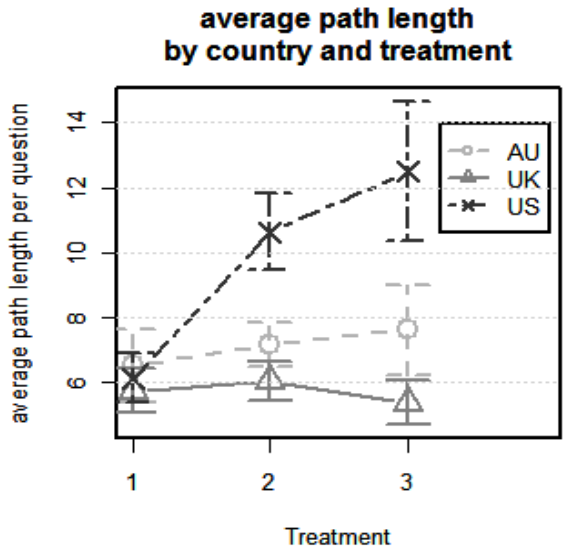

Figure 5c: Average path length of user to answer a question plotted by treatment and for each country with confidence intervals (0.95), including all questions that were accessed. Again, there is no significant difference for treatment 1. It is clearly observable that subjects needed many more clicks to find an answer on the US site. The UK does best followed by AU.

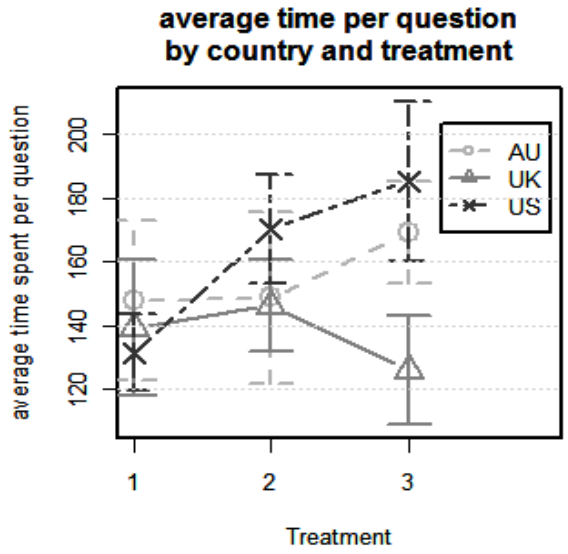

Figure 5d: Average time users spent per question plotted by treatment for each county, including confidence intervals (0.95). Again, the UK performs significantly better in treatment 3 than the other two countries. Note that in treatment 3 subjects on average spent 60s more on a question for US than for the UK.

Clearly the UK emerged from our experiments as the most 'navigable' of the sites; information was more quickly reached when users started at the home page. However, despite its poor performance on treatment 3, Australia's site performed equally well when users were free to use the internal search engine.

We also tested the extent to which the foreign offices competed well against other information providers, to test the hypothesis that governmental organisations face competition for nodality. For this comparison we used treatment 1, which simulated the normal situation of citizens being free to use external search engines. Faced with a particular information need, where would they turn to find the information? Table 10 shows the results specific to the first treatment.

Table 10: Using Unlimited Search (Treatment 1)

	AU	**UK**	**US**
external search used at least once	87%	100%	100%
average use of external search	61% (0.4)	75% (0.3)	80% (0.3)
average percentage of questions per user that were answered with foreign office site	70% (0.1)	53% (0.2)	58% (0.1)
average percentage of questions per user that were answered with government site from respective country	83% (0.1)	73% (0.1)	80% (0.1)
average percentage of questions per user that were answered with some government site (not necessarily from respective country)	93% (0.1)	84% (0.1)	90% (0.1)

First of all, our results show that the majority of subjects used a search engine. And second, their search led them only in about 60 per cent of the cases to the foreign office sites dedicated to provide this information. Differences between the UK and the US on the one hand, and AU on the other were particularly notable. Significantly more people used the AU foreign office site to answer the questions, indicating that the Australian site holds its own well against other information providers (although there may well be less such providers competing to provide information). The main competition seemed to be coming from other governments or other governmental organisations in the same country rather than from private information providers. In four-fifth of cases information was located on another government site of the respective country and if not, the sites of other governments, leaving only about 10 per cent of the information which was by non-governmental providers.

Discussion

To summarise our results we use the measures reported here to assess the online presence of the three foreign offices along the five dimensions of nodality discussed above (visibility, extroversion, accessibility, navigability and competitiveness). Visibility can be measured via the inlink analysis, while outlinks are used for measure of 'extroversion'. The percentage of external inlinks which go beyond the homepage into the site (so called 'deep links') is a measure of information accessibility. We have measured navigability in two ways, first via the webmetrics (SCC and path length analysis) and second the user experience measure via treatment 2 (users cannot use external search but can use the internal search engine) and 3 (no search allowed) in the experiment. Finally, competitiveness can be assessed using the extent to which questions in treatment 1 (subjects can use unlimited search) were answered using the site under observation.

	Visibility	**Extroversion**	**Accessibility**	**Navigability**	**Competitiveness**
Australia	Medium	Low	High	Medium/Low	High
UK	High	High	Low	High	Low
US	Low	Medium	Low	Low	Low

All of these characteristics represent some dimension of nodality. The table shows the UK site performing well across the majority of dimensions but less so in another two. Visibility is high according to the number of inlinks, which we know influence page rank algorithms of search engines (aided by the high number of outlinks which also feeds in to some extent). The site is by far the most 'extrovert' of the three, with a high number of outlinks to both private and public sectors. Its high levels of navigability from the path analysis of the webmetric section was borne out in the experiments, particularly treatment 3 where users started from the homepage and clicked their way through the site. With respect to accessibility however, the UK scored badly, with a high proportion of inlinks going to the home page rather than 'deep' into the site, suggesting that users looking for information held on the site will only arrive at the homepage. Although we have no real evidence that this is the case from our experiments, it is the case that the UK is the least competitive of the sites, with only just over half of responses using the foreign office site to provide the required information. The other providers are governmental organizations, largely from within the UK government, suggesting some weakness of the nodality of the foreign office vis a vis other departments.

The US was a middle ranking site on most of our dimensions apart from navigability. The site is huge, and some commentators would argue that such a large site could not hope to compete with the other two on navigability. However, the UK clearly outranked the similarly sized Australian site in both the webmetric and experimental analyses of navigability, so it is possible for structure to overcome the challenges of size under these metrics. In addition of course, it could well be argued that there is no need for the US site to be so large. In terms of visibility, the site only just outranked the UK in total numbers and fell way below even Australia when the indicator was normalized. The navigability measures calculated via path length showed the US not far behind the UK but the site dropped behind in treatments 2 and 3 in the experiments. The site clearly faces competition from other governments and other governmental sites in the US.

Australia's comparatively small site lagged behind the other two in terms of extroversion, being a prime example of an introverted web site. However, in terms of incoming links, once the figures were normalized its score for visibility was higher than that for the US. It scored badly in terms of navigability when users had no access to internal search but well in terms of accessibility, with the majority of links going 'deep' inside the site, suggesting that more content is readily available to users using external search. The poor navigability (in terms of the long path length) will work against this quality when users arrive via the home page. But the Australian site seems to fare well when competing against other governmental providers of the information we asked subjects to find in our experiments.

Conclusion

The proliferation of e-government initiatives and rankings to assess them across the world during the first half of the 2000s is probably the e-government equivalent of the e-commerce boom of the previous decade. Without wishing to denigrate the huge range of activity carried out during this period, we would like to suggest that the dimensions we have started to develop here might provide the basis of a more rigorous assessment of e-governments – and the implications of digital technologies for policy-making – in the future. We have focused entirely on information seeking, on the assumption that such interactions are by far the most common between citizens and governments. Commentators only started to wake up to the fact that the original expectations of the e-commerce boom were finally being realized when they focused on how different patterns of information seeking had changed people's buying habits (*The Economist*, 'E-Commerce takes off', 15th May, 2004). For e-commerce, the final 'closing of the deal' might or might not take place by

telephone or face-to-face, but the important shift is that virtually all the research is carried out on-line. The same may well be true for e-government.

We suggest that the methods we have used here - web metrics and user experiments - are well suited to study e-government. Until now there has been a tendency for webmetric analysis to be applied either to the narrow field of academic scientific communities – or to the internet as a whole. The huge investment that many states have made in their e-government initiatives, the size and importance of governments' online presence and the shift online in citizens' information seeking habits all suggest that the field could usefully be applied to the government domain. We have only looked at one department here, but hope that the methods we have used could be extended to other departments and ultimately 'whole of government' domains. Likewise, 'government on the web' lends itself to user experiments of a kind that could rarely be carried out for government before. Experimental research can be expensive and time consuming, but we hope that the methods we have used here could be extended, particularly in the development of ways to assess 'visibility' and 'accessibility' when subjects are using unlimited search. The sites we examined produced similar results in these conditions, but variations in the link structure suggest underlying differences in 'visibility' worthy of further investigation. Ultimately, experiments like these might be used to assess how citizens use the information they find on-line to make government-related decisions, such as choosing a school of university.

Governments seeking to develop their electronic interactions with citizens might do well to pay attention to these types of indicator of their performance. Clearly, governments have to plan for a variety of user behaviours when designing their on-line presence, as users will come via a variety of routes and will behave in different ways when they arrive at the site. But there is substantive research evidence to add to our own experimental findings that most users use external search engines (see for example Nielsen NetRating of 30th March 2006 which suggested that 84 per cent of the Internet population used a search engine in the month of January 2006) and will be disinclined to look further than the top search results or to navigate more than around a very limited number of clicks (3-6) once within a site(Bernardo et al, 1998). Therefore, it might be argued that visibility, and accessibility are the most important dimensions of nodality to prioritise.

In the on-line world there are more strategies than ever before available to governments aiming to maximize their nodality. At the same time there are new challenges, particularly the competition faced from other governments, private sector corporations and voluntary organizations. Governments that face a net loss of nodality on-line could be forced to turn to more costly tools in

terms of expenditure and policy instruments which cause more 'trouble, vexation and oppression' (Hood, 1983; Hood and Margetts, 2006) on their citizens.

Acknowledgements

This paper is based on the research project 'Participation in Internet-mediated Interactions', which formed part of the Communications Research Network (CRN), a Knowledge Integration Community funded by the Cambridge-MIT Institute and co-funded by British Telecom. The CRN brought together researchers from Cambridge University, MIT and University College London – economists, public policy experts, management analysts, engineers and computer scientists – to investigate all aspects of tomorrow's communications and computing technologies and their exploitation (see www.communicationsresearch.net). The project was based at the School of Public Policy, the Department of Computer Science and the Department of Economics, University College London (UCL). The principal investigators on the project were Professor Ingemar Cox (UCL Departments of Computer Science and Electrical and Electronic Engineering), Professor Helen Margetts (Oxford Internet Institute, University of Oxford) and Professor Steffen Huck (UCL Department of Economics) and the research fellows were Tobias Escher and Vaclav Petricek. The authors of this paper are very grateful to their co-investigator Professor Steffen Huck for his assistance in designing the user experiments and for the use of the ELSE laboratory. We would also like to thank Sally Welham (School of Public Policy) for administering the grant.

Contact Details:

Professor Helen Margetts

Oxford Internet Institute

University of Oxford

1 St Giles

Oxford

OX1 3JS
www.oii.ox.ac.uk, www.governmentontheweb.org

Tel: +44 (0)1865 287207
Email: helen.margetts@oii.ox.ac.uk

References

Accenture (2005) *Leadership in Customer Service: New Expectations, New Experiences*, The Government Executive Series, April 2005 (and reports in 2002-2004 available at www.accenture.com)

L. Adamic, B. Huberman (2001) 'The Web's hidden order' *Communications of the ACM*, vol. 44 no. 9, Sept 2001.

R. Albert, H. Jeong and Barabási (1999) 'A. Diameter of the World-Wide Web' *Nature*, 401 (September 1999), 130.

A. Bernardo, Huberman et al (1998) 'Strong Regularities in World Wide Web Surfing', *Science*

K. Bharat, B.-W. Chang, M. R. Henzinger, M. Ruhl (2001) 'Who Links to Whom: Mining Linkage between Web Sites', in *ICDM*, 2001, pp. 51–58.

S. Brin and L.Page (1998) 'The anatomy of a large-scale hypertextual Web search engine' *Computer Networks and ISDN Systems*, vol. 30 no. 1–7 (1998), 107–117.

A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R Stata and A. Tomkins 'Graph structure in the web: Experiments and models' 9th WWW 2000.

A. Caldas (2004) 'On the Web Structure and Digital Knowledge Bases' On-line and Off-line Connections in Science'. in: C.Hine (ed.) *New Infrastructures for Knowledge Production: Understanding e-Science* (Information Science Publishing: Hershey, PA).

S. Chakrabarti, B. Dom, D. Gibson, J. Kleinberg, P. Raghavan, S. Rajagopalan (1998) 'Automatic resource list compilation by analyzing hyperlink structure and associated text', in Proceedings of the 7th International World Wide Web Conference, 1998.

S. Chakrabarti, B. E. Dom, S. R. Kumar, P. Raghavan, S. Rajagopalan, A. Tomkins, D. Gibson, J. Kleinberg (1999) 'Mining the Web's Link Structure', *Computer*, vol. 32 no. 8 pp. 60–67, 1999.

Cap Gemini, *Online Availability of Public Services: How is Europe Progressing? Web-based survey on electronic public services, Report of the fifth measurement*, October 2004 (European Commission Directorate General for Information Society and Media), October 2004.

D. Dalziel *Government online: A multi-country study of egovernment usage* (World Association of Research Professionals, 2004).

C. Demchak, C. Friis & T.M. La Porte (2000) 'Webbing governance: National differences in constructing the public face', in G.D.Garson (ed.) *Handbook of Public Information Systems* (New York: Marcel Dekker Publishers)

M. Diligenti, F. Coetzee, S. Lawrence, C. L. Giles, M. Gori (2000) 'Focused Crawling using Context Graphs' in 26th International Conference on Very Large Databases, VLDB 2000, Cairo, Egypt, 10–14 September 2000, pp. 527–534.

W. Dutton, C. di Genarro and A. Millwood (2005) *The Internet in Britain: The Oxford Internet Survey* (OxIS), Hargrave, May 2005,

N. Eiron, K. S. Mccurley, 2003 'Locality, Hierarchy, and Bidirectionality in the Web'.

G. W. Flake, S. Lawrence, C. L. Giles, F. Coetzee (2002) 'Self-Organization of the Web and Identification of Communities' *IEEE Computer*, vol. 35 no. 3 pp. 66–71, 2002.

D. Gibson, J. M. Kleinberg, P. Raghavan (1998) 'Inferring Web Communities from Link Topology', in UK Conference on Hypertext, 1998, pp. 225–234.

I. Graafland-Essers, E.Ettedgui (2003) *Benchmarking EGovernment in Europe and the U.S.*

M. Hindman, K. Tsioutsiouliklis and J.A. Johnson (2003) Googlearchy: How a Few Heavily-Linked Sites Dominate Politics on the Web.

C. Hood (1983) *The Tools of Government* (London: Macmillan)

C. Hood and H. Margetts (2006) *The Tools of Government in the Digital Age* (London: Palgrave).

B. Huberman, L. Adamic (1999) 'Growth dynamics of the World-Wide Web', *Nature*, vol. 399 pp. 130.

B. A. Huberman, P. Pirolli, J.E. Pitkow and R.M. Lukose, R.M. (1998) 'Strong Regularities in World Wide Web Surfing', *Science*, 280 (April 1998), 95-97.
P. Ingversen (1998) 'The calculation of Web impact factors' *Journal of Documentation,* 4(2), 236-243.
International Telecommunication Union (2004) Internet indicators: Hosts, Users and Number of PCs.
J. M. Kleinberg ,'Authoritative sources in a hyperlinked environment', *Journal of the ACM*, vol. 46 no.5, 604–632, 1999.
R. Kumar, P. Raghavan, S. Rajagopalan, A. 'Trawling the Web for emerging cyber-communities'
J. Laherrre and D. Sornette (1998) 'Stretched exponential distributions in nature and economy: 'fat tails' with characteristic scales', *The European Physical Journal B - Condensed Matter*, 2(4):525—539.
T. La Porte, C. Demchak. & C. Friis (2001) 'Webbing Governance: Global Trends across National Level Public Agencies', *Communications of the ACM*, January
Linyuan Lu. (2000) Diameter of Random Massive Graphs in *Proceedings of the twelfth annual ACM-SIAM symposium on Discrete algorithms*.
W. Nooy, A. Mrvar and V. Batagelj (2005) *Exploratory Social Network Analysis* with Pajek. (Cambridge University Press).
L. Page, S. Brin, R. Motwani, T. Winograd (1998) 'The PageRank Citation Ranking: Bringing Order to the Web' Tech. Rep.,Stanford Digital Library Technologies Project.
H. W. Park (2003) Hyperlink Network Analysis: A New Method for the Study of Social Structure on the Web, *Connections* 25(1): 49-61
V. Petricek, T. Escher, I. Cox and H. Margetts (2006) *The Web Structure of E-Government - Developing a Methodology for Quantitative Evaluation***,** paper accepted for the E-Applications Track at the World Wide Web Conference, Edinburgh 23rd - 26th May 2006. see www.governmentontheweb for paper and presentation)
Taylor Nelson Sofres (2003) *Government Online: An international perspective*, Taylor Nelson Sofres.
M. Thelwall (2003) 'Conceptualizing documentation on the Web: an evaluation of different heuristic-based models for counting links between university web sites'
M. Thelwall. & D. Stuart (2006, to appear). 'Web crawling ethics revisited: Cost, privacy and denial of service'. *Journal of the American Society for Information Science and Technology.*
Tomkins, *Computer Networks* (Amsterdam, Netherlands: 1999), vol.31 no. 11–16 pp. 1481–1493, 1999.
United Nations (2003) *World Public Sector Report 2003: EGovernment at the Crossroads* Department of Economics and Social Affairs (New York: United Nations).
D. West (2005) *Digital Government: Technology and Public Sector Performance* (Princeton University Press).
H. XHe, C. Zha, H Ding, H. Simon, (2001) 'Web document clustering using hyperlink structures'.
S. Zhou and R. J. Mondragon (2004) 'Accurately modeling the Internet topology', *Physical Review* E, vol. 70, no. 066108, the American Physical Society, Dec. 2004.

Methodological Appendix.

This short appendix describes the key methods used for (1) the crawling of web sites to establish the structural metrics and (2) the user experiments we undertook to establish user metrics.

Webcrawling

We crawled all the sites (shown in table 1) and dumped the results in a format suitable for network metrics computation in Pajek. In addition we manually identified and blacklisted spider traps including the mailing list archive and diary application on the US site. The crawler was able to extract links from HTML documents but not from other content (for example*.pdf, *.doc, or flash). The links omitted form a negligible fraction of links followed by users though.

In general crawling is a technically complicated task that requires constant monitoring. We used the opensource crawler Nutch version 0.6. We had to modify it to use HTTP/1.1 as the Lotus Domino servers, very popular with governments, incorrectly handle zero content-length HTTP/1.0 requests and return HTTP Error 400 "Bad Request", preventing Nutch from crawling the site completely. Nutch was configured to run a single thread with 5s delay between requests and to store all links including the internal ones. The PC used to perform the crawls was a Intel(R) Pentium(R) 4 CPU 2.80GHz with 1G RAM and over 1TB networked disk storage. Neither hardware nor software proved to be a bottleneck but we had to limit the rate of our queries in order to not turn our crawling into a Denial-of-Service attack.

Each crawl was started from the homepage of the respective site and limited to domains as described in Table 1. An exhaustive crawl was performed with a security limit on the depth of 18. Once collected, the crawls were dumped, and post-processed by perl scripts to convert them to a format suitable for analysis in Pajek.

(2) User Experiments

The experiment was conducted at the computer laboratory of the ESRC Centre for Economic Learning and Social Evolution (ELSE) based at UCL's Economics Department. The lab provides 25 Microsoft Windows PCs, all of which are connected to the Internet via a server and have access to a central file storage. Each computer was set up with the same profile to ensure the conditions were equal for all participants. We used the Open Source browser Firefox with the Slogger-Extension that

enabled us to log the pages accessed by subjects in the course of the experiment. We controlled which sites subjects could access by routing the Internet connection via a proxy server that could be set up to block particular sites.

The participants of our study consisted of a self-selected sample of 135 people from various backgrounds but most of them were students. The requirements for this study were basic computer skills. The majority of our participants were aged between 20 and 25 years and regular users of the Internet. Recruitment took place through the subject database of the ESRC Centre for Economic Learning and Social Evolution. The database consists of volunteers that have indicated that they would be available for experiments. Potential subjects were contacted via email and were allocated to one of the sessions on positive reply. Subjects were seated so that they could not collaborate and were constantly monitored by two researchers. At the beginning of each session participants received a verbal introduction into the purpose and organization of the study and the amount of compensation they could expect. Every subject received a flat rate of £5 for showing up and a further £0.50 for each of the 16 questions that would be answered correctly.

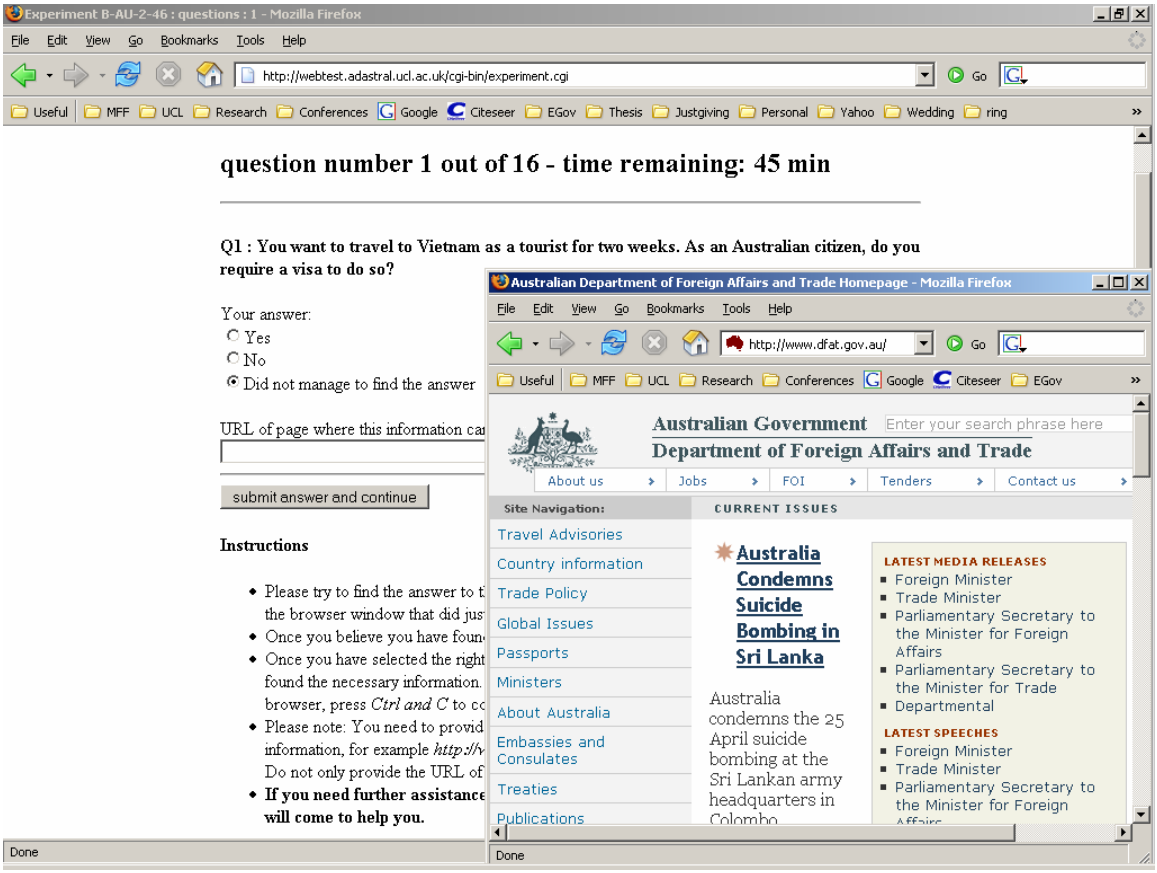

Figure 6. Screenshot of experiment interface, displaying one window with the online questionnaire and one window (resizable) that could be used to surf the Internet in order to find the relevant information for answering the question.

The participants had 45 minutes to answer all questions. The interface would inform them about the time they have taken so far, the total as well as the remaining number of questions. The performance-related payment was directed at motivating the subject to answer as many questions as possible. It was not allowed to return to a question once skipped. Participants were asked to select the correct answer plus to provide the URL of the page on which the information was found. At the end of the time limit or after all 16 questions had been answered, the participants were asked to give some information about their Internet usage and feedback on the experiment.