



# Australian Population Association

---

12th Biennial Conference

population and society: issues, research, policy

15-17 September 2004 - Canberra, Australia

## Australia's online presence: What potential migrants find out about Australia from the WWW

**Robert Ackland and Edith Gray** (robert.ackland@anu.edu.au)

The Australian National University

### Abstract

We provide a quantitative characterisation of the information that is available to potential migrants to Australia via the World Wide Web (WWW). Drawing on new methods adapted from the fields of web mining and data visualisation, we suggest that Australia's online presence, as encountered by potential migrants to Australia, is defined by a relatively small number of heavily-linked government and commercial sites. The tendency of organisations to use hyperlinks to cross-reference content on different websites has implications for the visibility of information targeted at different types of migrants. In particular, we show that the online information environment is mainly geared towards skilled and business migrants to the extent that other potential migrants (such as those wishing to enter Australia under *Family Stream*) could have difficulty locating relevant information on the WWW. We link our approach and findings to on-going research into the existence of the "digital divide" and conclude that research into this important issue must involve quantitative analysis of the accessibility and targeting of information in cyberspace.

Paper prepared for the 12<sup>th</sup> Biennial Conference of the Australian Population Association, 15-17 September 2004, Canberra.

# 1 Introduction

Up until very recently, a person considering migrating to Australia would have had access to at most two main sources of information specifically targeted at potential migrants (for example, about Australia as a place to live, and the process of migrating to Australia). The most obvious source of information would have been family or friends already living in Australia, while a second source would have been the Australian embassy or consulate in the person's home country, probably in the form of brochures or an interview with an immigration official.

Today, there is a third major source of information for potential migrants to Australia: the World Wide Web (WWW). The Internet, and in particular the WWW, can be used to provide information to prospective migrants in a much more diverse and dynamic fashion. Recent research into politics on the WWW by Hindman et al. (2003) and Ackland and Gibson (2004) has involved the use of methods from the fields of information science and statistics to assess the existence of online political networks and availability of political information on the WWW. We extend the methods described in Ackland and Gibson (2004) and present a new approach for evaluating the role of the WWW in the provision of information to prospective migrants to Australia.

We construct a "connectivity database" containing meta data on over 11,000 web pages that could be encountered by a person starting their search for information on migrating to Australia in the popular search engine Google, and then following hyperlinks to other pages. Our connectivity database stores meta data about the pages that can be collected using automatic methods and we focus our analysis on generic top-level domain (TLD) codes (e.g. .com, .edu) and country TLD codes (e.g. .au, .uk). We also collect information on the hyperlinks between different web pages in the connectivity database and this enables us to construct measures of web visibility and draw conclusions about the type of information potential immigrants are likely to encounter.

We provide a quantitative characterisation of the online information that is available to potential migrants to Australia and conclude that Australia's online presence is defined by a relatively small number of heavily-linked government and commercial sites. We show that the tendency of organisations to use hyperlinks to cross-reference content on different websites has implications for the visibility of information targeted at different types of migrants. In particular, the online information environment is mainly geared towards skilled and business migrants to the extent that other potential migrants (such as those wishing to enter Australia under *Family stream*) could have difficulty locating relevant information on the WWW. We link our approach and findings to on-going research into the existence of the "digital divide" and conclude that research into this important issue must involve quantitative analysis of the accessibility and targeting of information in cyberspace.

## 2 Background

In this section we provide background information on Australia's migration program and outline our approach for analysing information on the WWW pertaining to migration to Australia.

### 2.1 *Australia's migration program*

We focus our attention on potential migrants who are interested in permanent movement to Australia. These migrants can apply for permanent visas under a variety of schemes. The schemes include the Migration and Humanitarian Programs, both of which include a number of sub-categories. Other groups are eligible for permanent residency such as New Zealand citizens,

and children born overseas to Australian citizens, but these groups are not the subject of this paper.

In 2003 there were 93,914 settler arrivals to Australia (DIMIA, 2004). Europe (particularly the United Kingdom) is still the main region of origin contributing to just over one-fifth of settler arrivals (21.5 per cent). Other large contributors are Oceania (16.5 per cent), Southeast Asia (16.3 per cent) and Northeast Asia (11 per cent). There were 66,748 settler arrivals under the Migration (non-humanitarian) Program. Most arrived under the *Skill Stream* (38,504) with a further (28,066) arriving under the *Family Stream*. Under the Humanitarian Program there were 9,569 arrivals.

The *Skill Stream* is designed to attract migrants who can contribute to Australia's economic growth (DIMIA, 2004: 43), and consists of migrants who have particular occupational skills, outstanding talents or business skills. The categories included in the stream are: Skilled-Australian Linked, Regional Linked, Employer Nomination, Business Skills, Distinguished Talent and Independent. These potential migrants are highly-sought after individuals, and are probably equally demanded by other countries who supplement their labour supply with a migrant intake (McDonald & Kippen, 2001).

## **2.2 *Immigration information on the WWW: Visibility versus retrievability***

The Internet was once considered a space that was praised for its inclusiveness (Staeheli, et al. 2002). It is often viewed as a forum for communities and groups to have a voice without the normal constraints evident in other mass-communication methods (including cost and censorship). In political science research, the WWW has been identified as a source of low-cost "narrowcasting" of political information, and thus has the potential to influence the political system by shifting power toward non-mainstream players. The availability of these community and non-mainstream websites have been seen as a strength of the Internet—but also as a weakness—as the Internet can be used as a medium for discrimination as evidenced by the 'hate sites' which proliferate (Huijser, 2004).

Hindman et al. (2003) argue that early notions that new information technologies would create a "level playing field" and thus foster political equality have proven to be incorrect because of a confusion between the *retrievability* of a web page and its *visibility*. While in theory every web page is equally retrievable (as long as the server on which the page is being hosted is active), the visibility of a web page is something that is not equal, and it is a relative concept that is largely influenced by the number of inbound links to the page (more inbound links provide more traffic to a page).

So what types of sites do potential migrants encounter? It appears that although there are a number of racist Australian web sites, we question whether these are actually the types of sites that migrants encounter. Unless they are highly linked to by other websites, it is unlikely they these types of sites will be visible. In comparison, business organisations are highly visible on the Internet, and it is suggested that the Internet is a forum better suited to e-business than e-democracy (Sunstein, 2001). One investigation on the use of web-pages by or related to immigrants finds that of 89 sites only eight are constructed by immigrants themselves, and many of the pages are sponsored by government agencies or policy think tanks (Staeheli, et al. 2002). Certainly the content of most of the web-pages of that research is 'information provision', that is, the pages provide information about immigration topics like obtaining a visa, rights and procedures and naturalisation.

We question what the implications of the above are for an analysis of the provision of information to prospective migrants via the WWW. Is the online information environment pertaining to migration to Australia dominated by a relatively small number of heavily linked

sites? Is a potential migrant who using the web to find out information about Australia (the “consumer” of the information) going to be able to find this information easily, or is she going to be “herded” towards heavily-linked commercial sites providing migration services? Is an organisation that is wanting to provide information to prospective migrants (the “producer” of the information) going to be faced with having web pages that are retrievable, but not visible?

### 2.3 *Constructing an information space*

To answer these questions we first need to be able to collect data on the information environment encountered by prospective migrants looking for information online. To completely characterise the information environment, we would need some way of classifying the content of each and every web page encountered (or likely to be) by migrants. This is clearly not feasible in the context of a large-scale analysis. We propose that the information environment can be usefully characterised by looking at meta data associated with the web pages, rather than the web pages themselves.<sup>1</sup> Hence our study involves the construction of an *information space* (see, for example, Dodge and Kitchin, 2001) which contains meta data pertaining to web pages encountered by prospective migrants to Australia.

We construct our information space by adapting and extending a method introduced in Ackland (2004a) and Ackland and Gibson (2004). In particular, we use the UberLink software (Ackland, 2004b) to construct a “connectivity database” where the observations are web pages that could be encountered by potential migrants using the popular search engine Google to find out information about Australia in July 2004. Our database contains meta data on over 11,000 pages that could be encountered by a person starting their search for information in Google, and then following hyperlinks to other pages.

The variables (or fields) in the connectivity database are meta data about the pages that can be collected using automatic methods and we focus our analysis on the information contained in the domain names of the internet sites that are hosting the pages. Resources on the internet such as web sites are identified via unique numeric *IP (internet protocol) addresses* that consist of 4 numbers (between 0 and 255) separated by dots. The Domain Name System (DNS) translates easier-to-remember character-based *domain names* into IP addresses (for example, the domain name “www.example.com” might translate to 198.105.232.4).<sup>2</sup> Each domain name consists of a series of character strings (“labels”), separated by dots, with the rightmost label in a domain being referred to as its “top-level domain” (TLD). There are several types of TLDs within the DNS - we focus here on what are known as “generic” TLDs or “gTLDs” (e.g. .com, .edu) and “country code” TLDs or “ccTLDs” (e.g. .au, .uk). The ‘.com’ domain is intended for commercial entities (that is, companies), ‘.gov’ is used by government agencies; ‘.edu’ is reserved for educational facilities; ‘.net’ is used by many types of organizations and individuals globally but was historically intended for use by internet service providers; and ‘.org’ is intended for use by the non-commercial or non-government sector.

The other important meta data that we collect (via UberLink) is information that is derived from the content of the web pages. In particular, we collect information on the hyperlinks between different web pages in the connectivity database. We know for a given page *i* in our database,

---

<sup>1</sup> Meta data is literally “data about data”. Thus, our connectivity database does not contain the *content* of the web pages but information *about* the web pages.

<sup>2</sup> The DNS is managed and coordinated by the Internet Corporation for Assigned Names and Numbers (ICANN). See <http://www.icann.org/> for more details.

what other pages page  $i$  links to and what pages (in our database) link to page  $i$ . We thus are thus able to construct a *web graph*.<sup>3</sup>

Having constructed our information environment, we then turn to characterising or analysing the online information space encountered by potential migrants. It is our contention that even though we are only collecting meta data pertaining to web pages, our data can be used to answer important questions about the provision of information to prospective migrants to Australia. The web graph data will allow us to draw conclusions about the visibility of different pages in the information space. In combination with the TLD codes we are able to determine *who* is providing information to prospective migrants and this in turn provides insights into *what information* is being provided and therefore, *what types of potential migrants* are being catered for on the WWW.

### 3 Constructing the connectivity database

In this section, we describe the construction of the connectivity database that is later used to study the online information space encountered by potential migrants to Australia. As discussed in Ackland (2004a), the construction of the connectivity database is analogous to constructing a snowball sample (see, for example, Frank and Snijders, 1994). First, the “initial sample” or “seed set” is selected, and then a web crawler is used to find the pages that pages in the seed set link to, thus forming the first wave of the snowball sample, or the “1<sup>st</sup> ring set”. Additional waves or ring sets can be constructed in a similar manner.

#### 3.1 The seed set

The method for constructing the seed set or initial sample of web pages depends on the research question. Ackland (2004a) and Ackland and Gibson (2004) study the networking behaviour of political parties and use seed sets consisting of homepages of political parties. In the present study, however, our seed set consists of pages that potential migrants would have encountered if they had used the popular search engine Google to search for information in July 2004.<sup>4</sup> Two common methods people use to do electronic searches are to enter a phrase relevant to the search, and to put in the most obvious keywords. In this case we chose two key terms: a phrase, ‘migration to Australia’; and two separate keywords, ‘migration’ and ‘Australia’. In what follows we refer to the former as the ‘migration to Australia’ search, and the latter as the ‘migration & Australia’ search.

For each search we compiled the resulting top 60 web-pages returned by Google. We then categorised each page according to the type of organisation it represented, using the following eight organisational types: government agency; migration agent; embassy, consulate or high commission; industry association; personal home page; education or research facility; other commercial organisation. We then excluded certain pages for two reasons. First, we excluded pages that were related to education or research on the rationale that these web-pages were not of interest, and would most likely not be browsed, by a potential migrant searching for information

---

<sup>3</sup> A *graph* consists of a set of vertices or nodes (representing, for example, people) and edges or arcs connecting the nodes (representing, for example, relationships between people). In a *directional graph*, the direction of an edge connecting two nodes is important, for example person  $i$  may have heard of person  $j$ , but not vice-versa, and hence there will be a single directional edge from node  $i$  to node  $j$ . The WWW can be modelled as a directional graph, with web pages represented as nodes and hyperlinks represented as directional edges.

<sup>4</sup> This method of constructing a seed set is similar to that used by Hindman et al. (2003) who propose a new quantitative methodology for studying political websites that also iteratively uses web crawlers to extract hyperlink information. The methods presented here have been developed independently to that of Hindman et al. (2003) and also differ in several key respects.

on migrating to Australia. Second, we dropped any page if the organisation represented by the page was already in the sample. For example, [www.how2immigrate.net/australia/permanent-visa.html](http://www.how2immigrate.net/australia/permanent-visa.html) was ranked 12 in one of the searches and [www.how2immigrate.net/australia/](http://www.how2immigrate.net/australia/) was ranked 46. We excluded the page ranked 46 because the organisation represented by the page (a migration agent) was already in the sample. These exclusions left 32 pages from the ‘migration to Australia’ search and 40 pages from the ‘migration & Australia’ search.

We then compiled our final list of seed sites by integrating the two searches based on the page ranking. The final list is in rank order of the two lists. We start by taking the number one page from each list. We then determined where that page was ranked on the other list, and took the one that had the highest ranking on the second list. So for example, the two number one pages were: [www.immi.gov.au/](http://www.immi.gov.au/) (No.1 on ‘migration to Australia’, and No. 3 on ‘migration & Australia’), and [www.migrationint.com.au/](http://www.migrationint.com.au/) (No. 6 on ‘migration to Australia’ and No. 1 on ‘migration & Australia’). Hence, [www.immi.gov.au/](http://www.immi.gov.au/) was ranked number one in our seed set, and [www.migrationint.com.au/](http://www.migrationint.com.au/) was ranked number two. We continued this throughout the two lists and ended up with a seed set of 50 ranked pages.

Table 1 provides the top 10 pages in our seed set. While the Department of Immigration (and Multicultural and Indigenous Affairs), DIMIA, homepage is the number one ranked site, it is striking that eight out of the top 10 web pages advertise the services of migration agents. As discussed further below, this suggests that Sunstein’s (2001) conclusion that the Internet is better suited to e-business rather than e-democracy may have parallels in the information space encountered by potential migrants to Australia. None of the top-10 sites are personal homepages set up to provide information for immigrants to Australia. Instead these pages are about institutions providing information about migration (such as visa provision and types) and about business providing information on services to assist in a movement to Australia. The sites most likely to be encountered provide information and services.

[Table 1 about here]

### 3.2 *The ring sets*

In order to construct an information space that can be used to study the online provision of information to potential migrants to Australia, we need to go beyond the pages returned by Google (which form the seed set). We cannot assume that potential migrants will simply look at the pages returned by Google; it is highly likely that they will follow hyperlinks to other pages (and often, to other organisations) in search of the information that they require.

To reflect this behaviour our connectivity database contains the seed set and two additional sets of pages that a potential migrant could encounter by following links from the seed set. The first set of pages is termed the “1<sup>st</sup> ring” or the “1<sup>st</sup> wave” of the snowball sample (see Ackland, 2004a for more details). UberLink was used to send a web robot or crawler<sup>5</sup> into each of the pages in the seed set and the 1<sup>st</sup> ring set is the set of pages returned from the web robot, with each page satisfying two criteria: (1) each page in the 1<sup>st</sup> ring set not also represented in the seed set (i.e. it must be a “new” page), (2) each page *i* in the 1<sup>st</sup> ring set must be “non-intrinsic” (i.e. not share the same domain name) to the page in the seed set that links to page *i*.

A few points are worth noting. First, criterion 1 ensures that each page in the connectivity database is unique. Second, the web crawler will follow intrinsic links (or links internal to the same website or domain) on a page in the seed set (up to a pre-specified number of links), but, as

---

<sup>5</sup> According to The Web Robots Pages, <http://www.robotstxt.org/wc/faq.html#what>, (authored by M. Koster, accessed 12/08/2004), a web robot is “... a program that automatically traverses the Web's hypertext structure by retrieving a document, and recursively retrieving all documents that are referenced.”

stated in criterion 2 above, only non-intrinsic links will be stored in the connectivity database. Without criterion 2, the connectivity database would be dominated by the pages of very large websites. Third, it should be remembered that a direct link between a page in the seed set and a page in the 1<sup>st</sup> ring set may in fact represent more than a single “step” or “jump” for a person following hyperlinks because the person may have to first follow intrinsic links within the website hosting the seed page before being taken “out” of the website to the page that is stored in the 1<sup>st</sup> ring set. Fourth, criterion 2 ensures that only links that are non-intrinsic to the page in the seed set currently being crawled are put into the 1<sup>st</sup> ring set. Of course, a given page  $i$  in the 1<sup>st</sup> ring set may be intrinsic to another page in the seed set or, indeed, any page in the connectivity database other than the page that “returned” page  $i$ . In fact, as discussed below, a lot of pages in the connectivity database share domains with other pages in the database.

The 2<sup>nd</sup> ring is constructed using a method analogous to the construction of the 1<sup>st</sup> ring.

The ring sets represent the availability of information to prospective migrants. Relevant pages in Google are easily available (just one click away), and are thus placed in the seed set. However a new organisation that is found by following links from a seed page is harder to get to (possibly several clicks away, depending on whether the person needs to first follow intrinsic links), and also possibly less relevant because of the problem of *topic drift*. We represent this by placing this new page into the 1<sup>st</sup> ring set. Similarly, a page that is found by crawling a page in the 1<sup>st</sup> ring set is even less accessible relative to pages in the seed set, and hence this page is found in the 2<sup>nd</sup> ring set. Note the assumption behind this is that prospective migrants are using a search engine such as Google as the “starting point” for their search for information; we feel that this assumption is quite reasonable.

### **3.3 Pages and page groups**

The connectivity database contains 11,906 observations, with each observation representing a unique web page. However, the analysis we want to conduct is at the level of the organisation or functional grouping rather than the web page. That is, we are more interested in constructing web graphs and quantitative measures derived from these graphs where, for example, the migration agent Dolphin Migration is represented as a single node in the graph, rather than a number of nodes with each node representing a different page in the Dolphin Migration website ([www.dolphinmigration.com.au](http://www.dolphinmigration.com.au)).

UberLink has a facility for aggregating web pages that come from the same organisation or functional grouping within an organisation (see Ackland, 2004b for more details). This process of creating *page groups* involves the analyst looking through the pages in the connectivity database and deciding on how to aggregate the pages. In most cases, the decision will be simply to place all pages with the same domain name into the same page group. For example, in our database we have four pages from the website of the “Victoria online” website of the Victoria government: [www.vic.gov.au](http://www.vic.gov.au), [www.vic.gov.au/stategov/copyright.html](http://www.vic.gov.au/stategov/copyright.html), [www.vic.gov.au/stategov/disclaim.html](http://www.vic.gov.au/stategov/disclaim.html), [www.vic.gov.au/treasury/treasury.html](http://www.vic.gov.au/treasury/treasury.html). We aggregated these into a single page group, [www.vic.gov.au](http://www.vic.gov.au). However, in some situations it is useful to distinguish between different functional groups within the same organisation and therefore to create several page groups containing the pages from the same web site. For example, our database contains 159 pages from the DIMIA website but rather than aggregating these pages into a single page group, we used 13 separate page groups that represent different activities or functional groups within DIMIA, including [www.immi.gov.au/migration/family](http://www.immi.gov.au/migration/family) and [www.immi.gov.au/migration/skilled](http://www.immi.gov.au/migration/skilled).

### 3.4 Structure of the connectivity database

Our connectivity database contains 7,755 page groups (or “sites”): 50 in the seed set, 1,142 in the 1<sup>st</sup> ring, and 6,563 in the 2<sup>nd</sup> ring. Using page groups as the unit of analysis results in a 35 per cent decrease in the number of observations (compared with the number of pages held in the connectivity database) and thus a significant reduction in the scale of the data. It should be noted that where a page group contains pages from different rings, the group will be assigned to the ring closest to the seed set. For example, a page group that contains pages from the seed set and also the 1<sup>st</sup> and 2<sup>nd</sup> ring set will be allocated to the seed set.

The structure of the connectivity database is represented in Figure 1. Even though the connectivity database was constructed with two only iterations of the web crawler, we find that the depth of outbound links from a given page can be greater than two. For example, as shown in Figure 1, DIMIA links to the Migration Agents Registration Authority ([www.themara.com.au](http://www.themara.com.au)), which links to the Migration Institute of Australia ([www.mia.org.au](http://www.mia.org.au)), which links to Amnesty International Australia ([www.amnesty.org.au](http://www.amnesty.org.au)), which links to Amnesty International in Germany ([www.amnesty.de](http://www.amnesty.de)).

[Figure 1 about here]

UberLink has the facility to present interactive cybermaps using the 3D hyperbolic viewer H3Viewer of Munzner (1997). In Figure 2 a series of screenshots of H3Viewer cybermaps is presented showing the navigation through a web graph with the DIMIA website as the root node (or head of the graph). Moving from right to left in the cybermap for DIMIA shows the shortest path from DIMIA to the other page groups in the connectivity database.<sup>6</sup> If a page group is not present in the cybermap this means that the DIMIA site does not connect to that page group either directly or indirectly – there are 7388 nodes in the DIMIA web graph indicating that this site is connected (either directly or indirectly) to just about every other page group in the database. In the first screenshot, the link from DIMIA to [www.themara.com.au](http://www.themara.com.au) is shown – by clicking on the node representing [www.themara.com.au](http://www.themara.com.au), the focus of the graph is shifted (see the second screenshot), and the links from this site are revealed. The path from DIMIA to [www.amnesty.de](http://www.amnesty.de) shown in Figure 1 is shown in the cybermaps in Figure 2.

[Figure 2 about here]

The distribution of hyperlinks has been the subject of much research, with authors such as Barabasi and Albert (1999) and Kumar et al. (1999) finding that the distribution of both inbound and outbound links follows a power law.<sup>7</sup> This implies a very unequal distribution of links, with a small number of popular sites receiving a large proportion of the total links and less successful sites (e.g. personal homepages) receiving very few links. Hindman et al. (2003) argue that the distribution of inbound links has important implications for the visibility of political information on the WWW, and they find evidence to support the conclusion that inbound links in the political web communities that they study follow a power law distribution. It is of interest to look at the distribution of inbound links in the connectivity database used in the present study. In Figure 3, the (log of the) number of inlinks into a given site in the connectivity database is plotted on the horizontal axis and the (log of the) number of sites which have at least as many inlinks as that given site. The figure displays that, as has been found for the WWW in general, inbound links within the connectivity database are very unequally distributed. One site (DIMIA) receives a total of 67 inbound links from other sites within the connectivity database, while the next most highest inlinked site (the Department of Foreign Affairs and Trade) receives 36

---

<sup>6</sup>UberLink also has the facility to draw H3Viewer cybermaps showing inbound links to a particular site.

<sup>7</sup> See Hindman et al. (2003) for more details.

inbound links. At the other end of the spectrum we find that 710 sites receive only 2 inbound links and 6,416 sites (over 80 percent of the connectivity database) receive only a single inbound link.

[Figure 3 about here]

The data in Figure 3 form a straight line (with a bulge around the middle) – this is evidence that the inbound links in the connectivity database do follow a power law distribution. A linear regression using the data in plotted in Figure 3 produced an estimated slope coefficient of  $-0.45$  and an  $R^2$  of  $0.93$  – the relatively high  $R^2$  indicates that the data fit the linear regression line quite well. The slope coefficient is quite a bit smaller (in absolute terms) to what was found by Hindman et al. (2003) – the slope coefficients they estimated for 12 political web communities ranged between  $-1.252$  and  $-1.958$ , thus indicating an even more unequal distribution of inbound links that we have found here.

In general, our plot in Figure 3 looks quite different to Figures 1-3 in Hindman et al. (2003), but there are two reasons for this difference. First, Hindman et al. (2003) worked with much larger datasets containing approximately 250,000 pages (compared with our database of around 11,000 pages) – the larger scale of their database has implications for the number of inbound links to any given site (their highest-linked site typically had over 10,000 inbound links). Second, as discussed later, Hindman et al. have used sophisticated automatic methods to minimise the problem of topic drift and they investigate the distribution of inbound links only taking into account those sites that contain at least one “positive” page (i.e. a page that is relevant to a particular political issue). The average number of sites in the web communities studied by Hindman et al. is around 1,000 – much less than the approximately 7,755 sites we consider here (and we acknowledge that a large proportion of these sites are not relevant to migration to Australia, and have only been included in the connectivity database because of the problem of topic drift). In future work, we will use methods to minimise topic drift and we expect that will lead to the distribution of inbound links in our constructed connectivity database being more similar to that found by Hindman et al. (2003).

## 4 Visibility of information on the WWW

In this section, we use our constructed information space to provide a quantitative characterisation of the information that is available online to prospective migrants to Australia. The main aim is to assess the accessibility or visibility of information on the WWW, and we do this in two ways: a compositional analysis of the seed set and ring sets, and by calculating more general measures of web visibility using Kleinberg's (1999) Hyperlink-Induced Topic Search (HITS) algorithm.

### 4.1 Composition of seed set and ring sets

Conceptualising the constructed information space as a core (the seed set) surrounded by two rings is useful for two reasons. First, this structure represents how the database was actually constructed (in a manner analogous to the construction of a snowball sample). Second, the ring structure is suggestive of the availability or visibility of online information for prospective migrants to Australia. Assuming a person starts his or her search for information using a search engine such as Google, the pages in the seed set are going to be the most visible (and pages ranked higher by Google are going to be more visible than those ranked lower).<sup>8</sup> The page

---

<sup>8</sup> Note that we recognise that not all prospective migrants would start their search for information using a search engine; some may go directly to a website (such as DIMIA or a migration agent) that they are already aware of. We take this into account in other measures of visibility presented below.

groups in the 1<sup>st</sup> ring are going to be less accessible than those in the seed set, but still quite visible since each page in the 1<sup>st</sup> ring is by definition at most 1 step or degree of separation from at least one page in the seed set.<sup>9</sup> Similarly, the pages groups in the 2<sup>nd</sup> ring set will be even less visible.

In this section, we present a compositional analysis of the data in the seed set and two rings in our connectivity database.

#### 4.1.1 Composition of seed set

We have three types of meta data pertaining to the web pages in the seed set. First, we have categorised each page according to the eight organisational types discussed above. Second we have the meta data that is contained in the domain names: the generic TLD and the country-code TLD.

From Table 2, it is apparent that migration agents dominate the seed set (as was found for the top 10 ranked pages in the seed set). Migration agents make up 48 per cent of all pages, and other commercial pages (like migration lawyers) represent 14 per cent of the seed set. Although the number one ranked page is the DIMIA homepage, government web pages only make up eight per cent of the seed set, or 18 per cent if embassies, consulates and high commissions are included. Only a small percentage of pages are personal homepages (six per cent), suggesting that the visibility of community- or locally-sourced information is low at this initial level of page accessibility.

[Table 2 about here]

Pages in the seed set are most likely to be hosted from within Australia. The country code for the seed set indicates that almost half (48 per cent) are Australian sites (Table 3). A further 44 per cent do not have a country code identified. This is not surprising, given that many hostnames contain the generic TLD but no country code (i.e. the page ends in .com or .org for example). The remainder (8 percent) come from other countries including Austria, Thailand, Romania and Hong Kong.

By far the most prominent generic TLD is ‘.com’ with 72 per cent of seed pages being hosted on web sites using this domain. As discussed earlier, this reflects the prominence of migration agents and other commercial organisations. Ten per cent of generic TLDs are ‘.org’ and only eight per cent are government websites. The composition of the seed set, which is the most visible, looks quite different to the composition of the 1st and 2nd ring sets.

[Table 3 about here]

#### 4.1.2 Composition of 1<sup>st</sup> and 2<sup>nd</sup> rings

Unlike the seed set, we are not able to breakdown the 1st and 2nd rings by organisational type. Instead we only have TLD information for the page groups in the 1<sup>st</sup> and 2<sup>nd</sup> rings. While it would be of great interest to categorise the page groups in the rings using the organisational types used in Table 2, this would entail looking at over 7,000 websites (to gain an understanding of the organisation represented by the page). This is beyond the scope of the present project, and the scale of such a task points to the potential usefulness of a method of drawing representative samples of pages from a connectivity database.<sup>10</sup>

---

<sup>9</sup> As stated earlier, a person may need to click on more than 1 hyperlinks to reach a page in the 1<sup>st</sup> ring set if he or she first needs to follow intrinsic (internal) links from the page in the seed set before being taken “out” of the website.

<sup>10</sup> See Ackland (2004a) for a discussion of approaches for sampling data derived from the WWW.

As for the seed set, we provide details about the meta data of the 1st and 2nd rings. Table 4 provides the country of the hosting server. A substantial proportion of both the 1st and 2nd rings are hosted in an unknown country (42 and 40 per cent, respectively). There is also a large decline in the percentage of sites that originate in Australia. In the seed set 48 per cent of sites are Australian sites. This falls to 43 per cent in the 1st ring and 38 per cent in the second ring.

It is also noticeable that the two rings have a similar make-up. There is a slight increase in sites hosted in Canada and New Zealand between the 1st and 2nd rings—which reflects the corresponding decrease in sites hosted in Australia, but the difference between the two rings is marginal.

[Table 4 about here]

We also examine the generic TLD data for the 1<sup>st</sup> and 2<sup>nd</sup> rings (Table 5). As speculated, commercial websites predominate in this dataset. Over half of both rings are commercial websites indicated by the TLD ‘.com’, although this is a substantial decrease from what was found for the seed set, which have 72 per cent indicated by ‘.com’. In total, 53 per cent of the entire dataset are commercial websites supporting our proposition that what potential migrants are mainly exposed to is commercial interests. There is an increase between the seed set and the 1st ring for government webpages (from eight per cent to 17 per cent), and then a decrease in between the 1st and 2nd rings, although the decrease is only small (around 3 per cent). The proportion of “unknown” generic TLDs increases between the seed set, 1st ring and 2nd ring.

[Table 5 about here]

## 4.2 Authority/Hub scores

The above analysis of the composition of the seed set and ring sets gives some insights into the visibility of information on the WWW, but there are two limitations. First, the compositional analysis is valid (in terms of implications for visibility) only if the underlying assumption that prospective migrants use Google (and the pages they follow on to) to find their information is correct. But for many prospective migrants the starting point in the search for information may be, for example, the website of a migration agent or else a personal homepage that do not feature in the seed set. We want to get some idea of what information would be encountered by a potential migrant if they entered our web system *at any point*, rather than just via a website in the seed set.

A second problem with using the compositional analysis to draw conclusions about visibility is that it does not take into account the *number* of links to a given web site. It is intuitive that a site in the 1<sup>st</sup> ring that is linked to by several sites in the seed set be measured as having greater visibility than another site in the 1<sup>st</sup> ring that is linked to by only one site from the seed set. Hindman et al. (2003) argue that the number of inbound links to a site is a good predictor of web traffic to the site for three reasons: there are more online paths to heavily-linked sites, higher-quality sites tend to attract more links, search engines generate a large proportion of web traffic (and most modern search engine algorithms tend to return heavily-linked sites first).<sup>11</sup>

In order to make more general conclusions about web visibility in our web system, we make use of the information on inbound and outbound links within the system. In particular, we use Kleinberg's (1999) Hyperlink-Induced Topic Search (HITS) algorithm. This algorithm is based on the premise of the existence of two distinct, but inter-related, types of pages (or in our case, page groups). *Authorities* are highly-referenced pages on a particular topic, while *hubs* are pages

---

<sup>11</sup> Note that it is clear from this that the number of inbound links does not necessarily *cause* higher levels of traffic to the site, but is positively correlated with traffic levels.

that point to the authorities (and thus confer authority). There is a mutually reinforcing relationship between authorities and hubs: a good hub points to many good authorities, and a good authority is pointed to by many good hubs.

The HITS algorithm is implemented by associating each page group  $p$  in the connectivity database with an *authority weight*  $x(p)$  and a *hub weight*  $y(p)$ , which are initialised to 1. Let  $p \rightarrow q$  denote “page group  $p$  links to page group  $q$ ”. The HITS algorithm iteratively updates the  $x$ ’s and  $y$ ’s as follows:

$$x(p) = \text{SUM}_{\{q \rightarrow p\}} y(q)$$

$$y(p) = \text{SUM}_{\{p \rightarrow q\}} x(q)$$

Thus, in a single iteration,  $x(p)$  is replaced by the sum of the  $y(q)$ ’s of all page groups pointing to  $p$ , and  $y(p)$  is replaced by the sum of the  $x(q)$ ’s that page group  $p$  points to. After each iteration, the  $x$ ’s and  $y$ ’s are normalised so that the sum of squares is one. The HITS algorithm converges very quickly.

We ran the HITS algorithm over the connectivity database with the aim of studying the characteristics of the 20 most highly ranked (in terms of authority and hub weights) page groups. However we ran into an initial stumbling block – the top-20 rankings were dominated by page groups that had nothing to do with immigration to Australia. In particular, the top-20 visibility listings were dominated by sites such as: [www.hotelsinlasvegas-usa.com/](http://www.hotelsinlasvegas-usa.com/), [www.hotelsinvienna-austria.com/](http://www.hotelsinvienna-austria.com/) and [www.hotelsinnice-france.com/](http://www.hotelsinnice-france.com/) (there approximately 100 of these sites in the connectivity database and all feature in highly in terms of authority and hub scores). The similarity of the site names suggest that they are associated with a single organisation. They were brought into the connectivity database because they are connected to by a commercial web portal [www.clcdiscus.com](http://www.clcdiscus.com) that features in the seed set via the web page: [www.clcdiscus.com/Migration-Australia.html](http://www.clcdiscus.com/Migration-Australia.html) (this page was ranked 43 in our seed set). The outbound cybermap for [www.clcdiscus.com](http://www.clcdiscus.com) is shown in Figure 4.

[Figure 4 about here]

There are several observations that can be made. It is well-known that commercial web portals such as [www.clcdiscus.com](http://www.clcdiscus.com) use complicated hyperlinking schemes that are designed to raise the web visibility of affiliated sites and thus promote web traffic to these sites. The existence of [www.clcdiscus.com](http://www.clcdiscus.com) (and the sites it links to) in our connectivity database is testament to ability of commercial organisations to manipulate the “topology” of the WWW to suit their commercial interests. The web page [www.clcdiscus.com/Migration-Australia.html](http://www.clcdiscus.com/Migration-Australia.html) only appeared in our seed set because it was highly ranked by Google, and it is now apparent that this page was only highly ranked by Google because like most modern search engines, Google tends to return heavily linked sites first. The people behind [www.clcdiscus.com](http://www.clcdiscus.com) in a sense “fooled” Google into listing their site highly, even though it is only (at best) marginally relevant to a person looking for information on migration to Australia and is more highly ranked than other more worthy sites.

This is not meant to be an indictment of Google, since this is the company that has most successfully harnessed the topology of the WWW to provide a popular search engine service that is renowned for returning relevant information. Google is active in improving its technology to ensure that its results are not unduly manipulated by organisations. However, we point to this as a example of how organisations can manipulate of the topology of the WWW and this manipulation can have an impact on the type of information that people encounter on the WWW and their subsequent behaviour. While a potential migrant who looked at the [www.clcdiscus.com](http://www.clcdiscus.com) site may have quickly realised that the site was a commercial web portal and not likely to provide the information being sought, one can imagine other situations that might not be so “clear cut”.

The ability of organisations to use hyperlinks to raise the prominence or visibility of associated sites, thus forming discrete “web communities”, can have important implications for the conduct of both e-commerce and also politics on the WWW. By entering such a community it is possible that a visitor's attitude (towards a particular product or service, for example, or towards a particular political issue) may be reinforced or shaped.<sup>12</sup> See Ackland and Gibson (2004) for further discussion of this issue in the context of politics on the WWW.

Finally, we note that social scientists studying data from the WWW need to be aware of the problem of topic drift and the tendency for some commercial organisations to manipulate the topology of the web since these phenomena may have a significant impact on research findings. Further, there needs to be methods for identifying web communities that are irrelevant to the research question (such as the community in our connectivity database associated with [www.cldisuc.com](http://www.cldisuc.com)) and removing the offending observations so they do not skew the analysis.

To this end, the calculation of hub and authority scores via the HITS algorithm allowed us to identify the [www.cldiscus.com](http://www.cldiscus.com) web community which is irrelevant to our study and skewing the visibility results. We removed the web pages associated with this community and re-ran the HITS algorithm. In Table 6, we present the top-20 ranked page groups according to the authority and hub weights.

[Table 6 about here]

It is immediately apparent that government sites are the most visible in terms of both the authority and hub scores. While government sites account for less than 20 percent of page groups in the connectivity database, over half (12) of the top-20 authority sites are government sites and government sites make up nine of the top-20 hubs. Commercial sites are also well represented amongst the most visible sites, but their representation is less than would be expected given prominence in the connectivity database as a whole: while over 50 percent of the page groups in the connectivity database are commercial, they comprise less than 40 percent of the top-20 ranked hubs and authorities.

The prominence of government sites is not that surprising since government agencies have a tendency (possibly even policy?) of linking to one another. Government agencies are clearly influencing the web topology by linking to one another, but one can argue that this is a different phenomenon to what was found with [www.cldiscus.com](http://www.cldiscus.com) – the government agencies are simply promoting awareness of related agencies for the benefit of people visiting their sites, rather than trying to manipulate web rankings for commercial gain. However, it should be noted that our visibility rankings do appear to be influenced by the problem of topic drift – it is hard to see how some of the highly visible government agencies (such as the National Archives of Australia and the Australian Secret Intelligence Service) would be relevant to a prospective migrant to Australia. The same can be said for some of the highly-ranked commercial organisations such as the online newspapers. In future work we will address the problem of topic drift further by investigating manual solutions (e.g. using the UberLink interface to investigate instances of topic drift and exclude irrelevant pages from the connectivity database) and automatic solutions (e.g. using Bayesian techniques to automatically assess the relevance of a page according to its content).<sup>13</sup>

Australian sites dominate the list of the most-visible websites. While Australian sites only account for around 40 percent of our connectivity database, they comprise 90 percent of the top-

---

<sup>12</sup> Gibson, Lusoli and Ward (2002), for example, provide evidence that viewing websites can have a significant impact on the electoral preferences, particularly for younger voters.

<sup>13</sup> Hindman et al. (2003) investigate the use of Support Vector Machine (SVM) classifier algorithms to automatically classify web pages as either related or unrelated to the pages in the seed set.

20 hubs and authorities. This provides confirmation that Australia's online presence, as encountered by potential migrants to Australia, is clearly being established by sites hosted in Australia. Note that several of the migration agents in our seed set are international organisations that facilitate migration to other countries in addition to Australia. For this reason, it is entirely possible that a potential migrant to Australia could be exposed to web pages relating to migration to Canada, for example. It is further possible that a prospective skilled migrant who encounters web material relating to Canada while looking for information on Australia could be influenced into considering Canada as a potential home. However, the absence of any Canadian websites in the listing of most visible websites provides some objective quantitative evidence that Australia's online web presence is not being “encroached upon” by websites representing countries competing with Australia for skilled migrants.

A final point relates to the prominence of information aimed at skilled versus family migrants. Earlier we hypothesised that the WWW was most likely being used to promote Australia to prospective skilled migrants. While we need to more fully study the data, there is some initial support for this hypothesis: DIMIA's Business Skills Entry page ([www.immi.gov.au/migration/business](http://www.immi.gov.au/migration/business)) is ranked 16<sup>th</sup> highest in the authority listings, and DIMIA's Migrating as a Skilled Person page ([www.immi.gov.au/migration/skilled](http://www.immi.gov.au/migration/skilled)) is ranked 14<sup>th</sup> in the hub listings. The DIMIA Migrating as a Family Member page ([www.immi.gov.au/migration/family](http://www.immi.gov.au/migration/family)) does not appear in either of the listings. In addition, five of the top-20 ranked hub sites are commercial sites offering migration services to business/skilled migrants.

## 5 Discussion and Conclusions

This paper has presented an application and extension of concepts and methodologies currently under development in the context of politics and the web (see Ackland and Gibson, 2004) to evaluate the role of the WWW in providing information to potential migrants to Australia. We have used the research software UberLink (Ackland, 2004b) to construct an information space containing data on web pages that could potentially be encountered by prospective migrants to Australia looking for information on the WWW. We have used our constructed connectivity database to provide a quantitative characterisation of online information environment encountered by potential migrants. While our analysis is at a preliminary stage, we are able to offer the following key observations:

- Australia's online presence, as encountered by prospective migrants, is largely defined by a relatively small number of heavily-linked sites. This finding mirrors what has been found in studies of politics on the web (see, for example, Hindman et al., 2003). The most highly-visible (and thus more likely to be visited by prospective migrants) sites are run by government agencies and commercial entities.
- The information environment encountered by potential migrants appears to be heavily skewed towards skilled and business migrants.
- The sites encountered by prospective migrants are likely to be mainly ‘information’ and ‘service’ sites.

The finding that the information environment appears to be skewed towards skilled and business migrants requires further discussion. We believe that this phenomenon results from a perception, held by the organisations who are providing online information to prospective migrants, that potential migrants using the web to find information are mainly going to be applying under the skilled migration program. Why would this perception be held? It is possible (perhaps probable) that a potential migrant under the *Family Stream* would be more likely to gather information

from the family member already in Australia. Research on the “digital divide” (e.g. Hoffman and Novak, 2000) has shown that Internet adoption is significantly related to income, education and race; this further suggests that web-using potential migrants are unlikely to apply under the Humanitarian Program and are in fact most likely to be candidates for the *Skill Stream* of the Migration Program.

As discussed previously, skilled migrants are highly-sought after by Australia and other countries that supplement their labour supply with a migrant intake. Skilled and business migrants are also potentially valuable clients to migration service agencies who will help them in all aspects of moving to Australia. We believe that the WWW is being actively used by government and commercial organisations as a way of attracting skilled migrants to Australia. Australian federal government departments are using the WWW to compete for skilled migrants who may otherwise go to Canada or the US. State government departments are using the WWW to try to influence the location decision (within Australia) of skilled migrants – for example, whether to set up business in Melbourne, as opposed to Sydney. Commercial migration agents are using the WWW to win lucrative business associated with helping skilled and business migrants in the process of relocating to Australia.

In summary, we feel that our analysis suggests that the WWW being used to compete for skilled migrants, and we find that the “winner take all” phenomenon that has been observed in the context of politics on the web (see, e.g. Hindman et al. 2003) is evident in the information environment encountered by prospective migrants to Australia. The question is: is this a problem? We feel that the answer is “yes”, but the reason depends on who you are.

*Producers of information.* An organisation operating in this information environment, for example a migration agent wanting to attract clients or a government agency wanting to promote services to prospective migrants, must realise that on the WWW, retrievability does not equal visibility. The “build it and they will come” mentality has been shown to have shortcomings – a prospective migrant will not visit a website if they don't know that it exists. While we do not recommend the tactics to increase web visibility employed by commercial web portals such as [www.clediscus.com](http://www.clediscus.com), for example, it is clear that producers of information aimed at prospective migrants need to be aware of the implications of web topology (links) for accessibility and visibility of online information.

*Consumers of information.* The consumers of the information are the prospective migrants themselves - are they finding the information they want? Our analysis suggests that if a prospective *Family Stream* migrant from a poor country were to try to use the WWW to find out information about migrating to Australia, the chances are they would have a very difficult time locating what they want because Australia's online presence (as perceived by prospective migrants) appears to be catered towards businessmen from South Africa.

*Those concerned with equality of access to online information and the digital divide.* Previous quantitative research into the digital divide has focused on conducting surveys of web users and comparing the characteristics of the average web user with those of the population at large. Evidence for the existence of the digital divide has been presented in terms of significant differences in these characteristics e.g. men being more likely to use the web than women, or higher-income groups exhibiting greater web usage rates than the poor. The focus on user surveys has also led to recent claims such as the following by Cole (2004): “The so-called digital divide [in the US] is closing: the fastest growing populations of users are Latinos and African-Americans. Only 4 percent more men than women use the internet...”

We feel that while web usage surveys can give important insights into the existence or otherwise of a digital divide, it is impossible to adequately study the phenomenon without directly analysing the availability and targeting of information in cyberspace. The digital divide refers to

inequalities in the availability of online information to different segments of the population. The digital divide can occur because of differences in web access rates, but it also can occur because organisations are targeting online information to those groups from which they can expect the highest marginal return for their efforts (in the case of the present study, skilled/business migrants).

Some may believe that it is proper for the topology of the WWW to be largely determined by market forces, that is, for organisations to target the most valuable segments of the population, thus (indirectly) ensuring that online information of interest to other less valuable segments is less visible and therefore harder to find. However, there is every reason to expect that the WWW could be subject to “market failures” that need to be addressed by governments or authorities that wish to promote equality of access to information. Our research has proposed a method for assessing the existence of the digital divide by directly looking at the availability of information to different segments of the population.

## 6 References

Ackland, R. (2004a), “Using Adaptive Sampling Methods in the Analysis of Social Phenomena on the WWW,” paper presented at the 6<sup>th</sup> International Conference on Social Science Methodology, 17-20 August, Amsterdam, The Netherlands.

Ackland (2004b), “UberLink: Software for Analysing Networks on the WWW (User Guide),” mimeograph, The Australian National University.

Ackland, R. and R. Gibson (2004), “Mapping Political Party Networks on the WWW,” Australian Electronic Governance Conference, University of Melbourne, 14-15 April, 2004.

Barabasi, A. and R. Albert (1999), “Emergence of scaling in random networks,” *Science*, 286, pp. 509-512.

Cole, J. (2004), “On superhighway to tomorrow, today,” *The Australian Higher Education Section*, p. 44, 8<sup>th</sup> September 2004.

Department of Immigration and Multicultural and Indigenous Affairs (DIMIA) (2004). *Immigration Update 2002–2003*. Research and Statistics Division, DIMIA: Canberra.

Gibson, R. K., W. Lusoli and S. Ward (2002), “Online Campaigning in the U.K. : The Public Response.” Paper presented at the Annual Conference of the American Political Science Association, Boston.

Hindman, M., Tsioutsoulis, K. and J.A. Johnson (2003), “‘Googlearchy’: How a Few Heavily-Linked Sites Dominate Politics on the Web,” downloaded from <http://hindman.cc/>.

Huijser, H. (2004), “Internet hate: Exploring the limits of free speech,” *Australian Mosaic*, No.5: 27–28.

Kleinberg, J. (1999), “Authoritative Sources in a Hyperlinked Environment,” *Journal of the ACM*, 46(5), pp. 604-632.

Kumar, R., Prabhakar Raghavan, Sridhar Rajagopalan and A. Tomkins (1999), “Trawling the Web for emerging cyber-communities,” *Computer Networks (Amsterdam, Netherlands: 1999)*, 31, pp. 1481-1493.

McDonald, P. and Kippen, R. (2001), “Labor supply prospects in 16 developed countries, 2000–2050,” *Population and Development Review*, Vol.27(1):1–32.

Munzner, T. (1997), "H3: Laying Out Large Directed Graphs in 3D Hyperbolic Space," *Proceedings of the 1997 IEEE Symposium on Information Visualization*, October 20-21 1997, Phoenix, AZ, pp 2-10, 1997.

Staheli, L., Ledwith, V., Ormond, M., Reed, K., Sumpter, A. and D. Trudeau (2002), "Immigration, the internet, and spaces of politics," *Political Geography*, Vol.21: 989–1012.

Sunstein, C. (2001), *Republic.com*, Princeton University Press: Princeton.

Hoffman, D. and T. Novak (2000), "The Evolution of the Digital Divide: How Gaps in Internet Access May Affect Electronic Commerce," *Journal of Computer Mediated Communication*, 5.

Dodge, M. and R. Kitchin (2001), *Mapping Cyberspace*, London: Routledge.

**Table 1: The top 10 pages in the seed set<sup>a</sup> by Organisation type, July 2004**

Top 10 pages	Ranking	Organisation type
www.immi.gov.au/	1	Gov Department
www.migrationint.com.au/	2	Migration Agent
www.migrationaustralia.com.au/	3	Migration Agent
www.dolphinmigration.com.au/	4	Migration Agent
www.australia-migration.com/	5	Migration Agent
www.how2immigrate.net/australia/	6	Migration Agent
www.australian-embassy.at/migration.htm	7	Embassy
www.migrationexpert.com/	8	Migration Agent
www.meridien-migration.com.au/	9	Migration Agent
www.migrationbureau.com/australia/default.htm	10	Migration Agent

<sup>a</sup> Compiled from returns to two Google searches using first, the phrase ‘migration to Australia’ and second, the two separate keywords ‘migration’ ‘Australia’.

**Table 2: Organisation type of the 50 pages in the seed set<sup>a</sup>, July 2004**

Organisation type	Frequency	Per cent
Government Department	4	8
Australian Embassy, Consulate or High Commission	5	10
Migration Agent	24	48
Other commercial <sup>b</sup>	7	14
Personal homepage	3	6
Migration Industry Association	3	6
Other	4	8
Total	50	100

<sup>a</sup> Compiled from returns to two Google searches using first, the phrase ‘migration to Australia’ and second, the two separate keywords ‘migration’ ‘Australia’.

<sup>b</sup> ‘Other commercial’ includes Migration lawyers, homeloan services, and other business services.

**Table 3: Composition of seed set by country and generic TLD, %**

	%
<b>Country</b>	
Australia	48.0
Other	8.0
Unknown	44.0
Total N	50
<b>Generic</b>	
com	72.0
org	10.0
gov	8.0
net	4.0
Unknown	6.0
Total N	50

**Table 4: Composition of 1st and 2nd rings, country TLD, %**

	<b>1st ring</b>	<b>2nd ring</b>
Australia	42.6	38.4
United States	2.5	2.7
United Kingdom	2.5	2.8
Canada	0.9	2.9
New Zealand	0.7	2.2
Other	8.3	10.7
Unknown	42.4	40.3
Total N	1,142	6,563

**Table 5: Composition of 1st and 2nd rings, generic TLD, %**

	<b>1st ring</b>	<b>2nd ring</b>
com	54.7	52.0
gov	16.5	13.3
org	12.2	11.9
net	4.4	3.7
edu	3.1	5.5
Other	0.4	1.3
Unknown	8.8	12.3
Total N	1,142	6,563

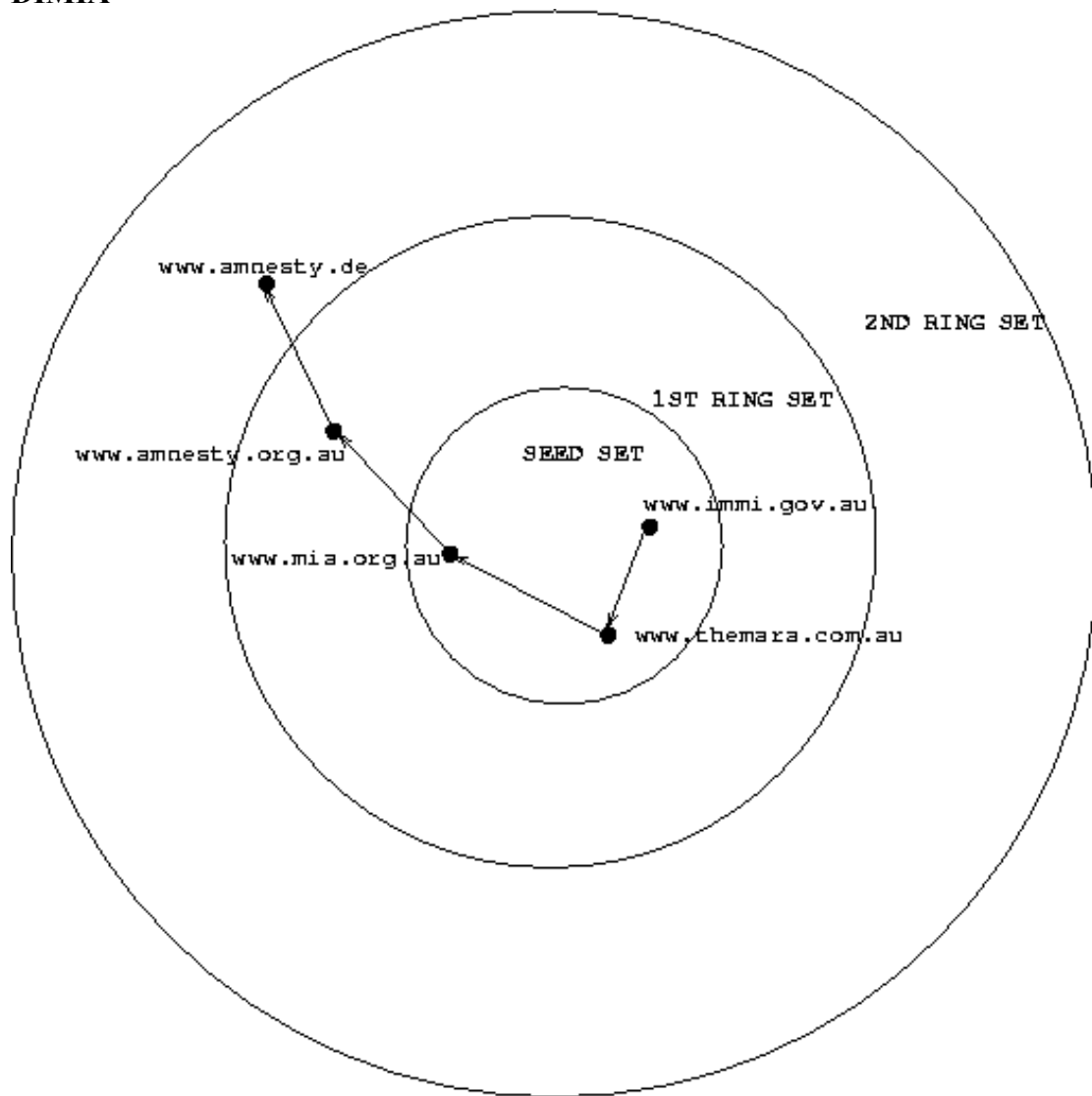
**Table 6a: Top 20 authority sites**

Rank	URL	Ring	auth.	hub	gTLD	ccTLD	Comment
1	<a href="http://www.dfat.gov.au/">http://www.dfat.gov.au/</a>	1	0.669	0.504	gov	au	contains news, travel advice etc
2	<a href="http://www.vic.gov.au/">http://www.vic.gov.au/</a>	1	0.387	0.026	gov	au	portal containing info and links on all things to do with Victoria
3	<a href="http://www.fed.gov.au/">http://www.fed.gov.au/</a>	1	0.270	0.000	gov	au	contains Australian government information and service links
4	<a href="http://www.lawaccess.nsw.gov.au/">http://www.lawaccess.nsw.gov.au/</a>	2	0.175	0.000	gov	au	site providing information about legal rights/concerns for citizens of NSW
5	<a href="http://www.newsphotos.com.au/">http://www.newsphotos.com.au/</a>	2	0.173	0.000	com	au	provides access to newsprint photos etc, like stock photo library
6	<a href="http://foxsports.news.com.au/">http://foxsports.news.com.au/</a>	2	0.173	0.000	com	au	online newspaper
7	<a href="http://www.advertiser.com.au/">http://www.advertiser.com.au/</a>	2	0.129	0.000	com	au	online newspaper
8	<a href="http://www.yamara.com/axe/">http://www.yamara.com/axe/</a>	2	0.129	0.000	com	?	message board/humour
9	<a href="http://www.naa.gov.au/">http://www.naa.gov.au/</a>	1	0.122	0.529	gov	au	national archives of australia
10	<a href="http://www.racv.com.au/">http://www.racv.com.au/</a>	1	0.122	0.000	com	au	automobile services
11	<a href="http://www.afp.gov.au/">http://www.afp.gov.au/</a>	2	0.108	0.000	gov	au	australian federal police services
12	<a href="http://volunteersearch.gov.au/">http://volunteersearch.gov.au/</a>	2	0.100	0.000	gov	au	volunteer search services
13	<a href="http://www.visionaustralia.org.au/">http://www.visionaustralia.org.au/</a>	2	0.096	0.000	org	au	providing services to visually impaired
14	<a href="http://www.asis.gov.au/">http://www.asis.gov.au/</a>	2	0.094	0.000	gov	au	australian secret intelligence service
15	<a href="http://www.apsc.gov.au/">http://www.apsc.gov.au/</a>	2	0.093	0.000	gov	au	public service commission
16	<a href="http://www.aceinfo.net.au/Services/NRS/">http://www.aceinfo.net.au/Services/NRS/</a>	2	0.093	0.000	net	au	services for hearing impaired
17	<a href="http://www.immi.gov.au/migration/business/">http://www.immi.gov.au/migration/business/</a>	1	0.092	0.001	gov	au	migrant visa info - business skills entry page
18	<a href="http://www.smartraveller.gov.au/">http://www.smartraveller.gov.au/</a>	1	0.092	0.067	gov	au	travel advice
19	<a href="http://www.angelfire.com/md/mjq/">http://www.angelfire.com/md/mjq/</a>	1	0.092	0.000	com	com	portal for web building services?
20	<a href="http://www.disr.gov.au/">http://www.disr.gov.au/</a>	1	0.092	0.007	gov	au	dept of industry and tourism - wanting to attract investment/business

**Table 6b: Top 20 hub sites**

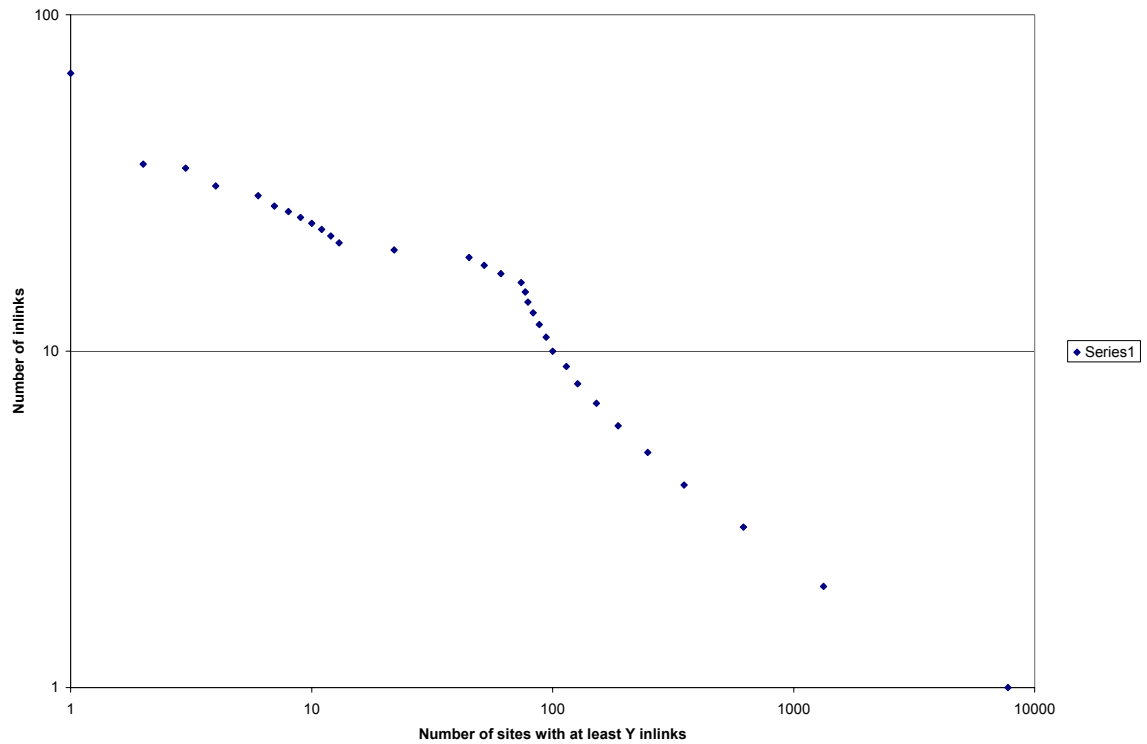
Rank	URL	Ring	auth.	hub	gTLD	ccTLD	Comment
1	<a href="http://www.naa.gov.au/">http://www.naa.gov.au/</a>	1	0.122	0.529	gov	au	national archives of australia
2	<a href="http://www.dfat.gov.au/">http://www.dfat.gov.au/</a>	1	0.669	0.504	gov	au	contains news, travel advice etc
3	<a href="http://mia.org.au/">http://mia.org.au/</a>	0	0.016	0.440	org	au	migration institute of australia - services provided
4	<a href="http://www.accesswa.au.com/">http://www.accesswa.au.com/</a>	0	0.000	0.390	com	?	migration services - western australia
5	<a href="http://www.theadvertiser.news.com.au/">http://www.theadvertiser.news.com.au/</a>	1	0.050	0.164	com	au	online newspaper
6	<a href="http://www.dewrsb.gov.au/">http://www.dewrsb.gov.au/</a>	1	0.062	0.127	gov	au	department of employment and workplace relations
7	<a href="http://www.comb.gov.au/">http://www.comb.gov.au/</a>	1	0.023	0.118	gov	au	commonwealth ombudsman wesbite
8	<a href="http://www.australian-embassy.at/">http://www.australian-embassy.at/</a>	0	0.000	0.116	?	at	australian embassy - vienna
9	<a href="http://www.lawfoundation.net.au/">http://www.lawfoundation.net.au/</a>	1	0.001	0.101	net	au	law foundation nsw
10	<a href="http://www.ausaid.gov.au/">http://www.ausaid.gov.au/</a>	1	0.018	0.096	gov	au	overseas aid programmes
11	<a href="http://www.anglia-migration.com.au/">http://www.anglia-migration.com.au/</a>	1	0.000	0.072	com	au	immigration consultancy privately run
12	<a href="http://www.australia-business-immigration.com.au/">http://www.australia-business-immigration.com.au/</a>	1	0.000	0.068	com	au	law firm in melbourne offering migration advice services
13	<a href="http://www.smartraveller.gov.au/">http://www.smartraveller.gov.au/</a>	1	0.092	0.067	gov	au	travel advice
14	<a href="http://www.news.com.au/">http://www.news.com.au/</a>	1	0.089	0.053	com	au	online newspaper
15	<a href="http://www.immi.gov.au/migration/skilled/">http://www.immi.gov.au/migration/skilled/</a>	1	0.009	0.038	gov	au	migration as a skilled person page
16	<a href="http://www.aasw.asn.au/">http://www.aasw.asn.au/</a>	1	0.001	0.033	asn	au	australian association of social workers
17	<a href="http://www.pm.gov.au/">http://www.pm.gov.au/</a>	1	0.024	0.033	gov	au	prime minister's home page
18	<a href="http://www.australianimmigrationadvisors.com.au/">http://www.australianimmigrationadvisors.com.au/</a>	1	0.000	0.031	com	au	providing migration services
19	<a href="http://www.aat.gov.au/">http://www.aat.gov.au/</a>	1	0.010	0.029	gov	au	administrative appeals tribunal
20	<a href="http://www.sydney.com.au/">http://www.sydney.com.au/</a>	1	0.002	0.027	com	au	portal - tourist info for sydney

**Figure 1: Structure of the connectivity database, and outbound path from DIMIA**





**Figure 3: Distribution of inbound links**



**Figure 4: H3Viewer cybermap for www.clcdiscus.com (outbound links)**

