WHAT CAN POTENTIAL MIGRANTS FIND OUT ABOUT AUSTRALIA FROM THE WWW?

Robert Ackland and Edith Gray

Printed in P*eople and Place*, vol. 13, no. 4, 2005, pp. 12-22

[Eyelid] The authors use a new approach to analyse the information that a prospective immigrant to Australia would be able to find out about the country from internet sites. They find that the most visible information is heavily slanted towards skilled or business immigrants and that most of it is provided either by governments or migration agents. Web sites hosted by community groups or individuals, or those slanted towards family reunion or humanitarian immigrants, are uncommon among the most visible sites. [End of eyelid]

INTRODUCTION

This paper investigates information about migration to Australia which is available on the World Wide Web (WWW). Up until very recently, a person considering migrating to Australia would have had access to two main sources of information specifically targeted at potential migrants. The most obvious source of information would have been family or friends already living in Australia, while a second source would have been Australian embassies or consulates in the person's home country, in the form of brochures or interviews with immigration officials. Today, there is a third major source of information for potential migrants to Australia: the World Wide Web. 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[1] involves the use of methods from the fields of information science and social science to assess the existence of online political networks and availability of political information on the WWW. We extend methods described in Ackland and Gibson[2] to investigate what information is available to prospective migrants via the WWW.

BACKGROUND

Australia's migration program

In this paper, we argue that there is a link between the visibility or availability of information on migration that can be found on the WWW and Australia's government policy towards immigration. It is important therefore to briefly consider Australia's migration intake.

Migrants to Australia 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.[3] In 2003 there were 93,914 settler arrivals to Australia.[4] 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.[5]

The Skill Stream is designed to attract migrants who can contribute to Australia's economic growth,[6] 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 also in demand by other countries that supplement their labour supply with a migrant intake.[7]

Given the level of skilled migration and the emphasis on it by Australian Government policies, we anticipate that information pertaining to skilled and business migration will be a central feature of the websites targeting potential migrants.

Immigration information on the WWW: visibility versus retrievability

Migration information on the WWW could relate to the process of migrating to Australia, but may also be about living in Australia, or aspects of Australia as a host country as experienced by previous migrants. Such information could even include anti-immigration sites. The wide variety of information available on the WWW stems from the ease with which individuals and organisations can create websites. The internet is often viewed as a forum for communities and groups to have a voice without the normal constraints evident in other mass-communication media (including cost and censorship) and it is a space that has been praised for its inclusiveness.[8] In political science research, the WWW is identified as a source of low-cost 'narrowcasting' of political information that has the potential to influence the political system by shifting power toward non-mainstream players. Community groups and non-mainstream organisations can put up websites with relative ease. The availability of such sites is both a strength of the internet and also a weakness as the internet can be used as a medium for discrimination as evidenced by the proliferation of 'hate sites'.[9]

Early research into the impact of new information and communication technologies suggested that improved accessibility of information via the WWW would create a 'level playing field' thus fostering political equality. However, while in theory every web page is equally *retrievable* (as long as the server hosting the page is active), the *visibility* of a web page is a relative concept that is largely influenced by the number of inbound links to the page.[10] Search engines such as Google tend to rank more highly those web pages that have many other pages linking to them.[11] Thus, while the information on community- or individual-run websites aimed at prospective migrants to Australia may be just as retrievable as the information on the website of a migration agent or government agency, there may be marked differences in the visibility of the different sources of information. Commercial or government pages may be ranked much more highly by search engines such as Google.[12]

We believe our study is the first to use new information retrieval methods to characterise the information available to migrants on the WWW in a quantitative fashion. Previous research using manual methods of data collection examined 89 sites by or related to immigrants. This found that only eight sites were constructed by immigrants themselves, and that many of the sites were sponsored by government agencies or policy think tanks.[13] The study highlighted the relatively low visibility of sites run by individuals and also found that the majority of web pages studies focused on 'procedural information' (for example, how to obtain a visa, applicants' rights, immigration procedures and naturalisation), that is, they focused on information on services provided by government and business. This finding reinforces other research which concludes that the internet is a forum better suited to e-business than e-democracy.[14]

As indicated, while there may be a wide variety of information on the WWW for potential migrants, we believe that certain types of information will be more visible. In particular, we feel that the most visible information will be that targeted toward skilled migrants. There are two reasons for this. First, skilled migrants are the largest group of migrants to Australia. Second, skilled migrants are valuable potential clients to migration agents and valuable potential citizens for the Federal and State governments. Thus the web is being used as a tool to 'compete' for these migrants.

Below we test this hypothesis empirically by investigating the visibility, and hence availability, of information targeted at different types of migrants to Australia.

DATA AND METHOD

The information environment encountered by prospective migrants to Australia can be usefully characterised by studying meta data associated with web pages, rather than the content of the web pages themselves.[15] This is the key feature of our study which distinguishes it from previous research in this area. We therefore construct an 'information space',[16] and contend that our methods of data collection and analysis are useful and appropriate. This is especially so when one considers the potential vastness and dynamism of the web which makes content analysis of individual web pages difficult or even infeasible.

We used new research software[17] to construct a 'connectivity database'. Here the observations are the web pages that could have been encountered by a potential migrant looking for information about Australia using the Google search engine (and then following hyperlinks to other pages) in July 2004. The fields in the connectivity database are meta data collected using automatic methods and we focus on generic top-level domain (TLD) codes (e.g. .com, .edu) and country TLD codes (e.g. .au, .uk).[18] We also know for a given page i in our database, what other pages page i links to (via hyperlinks) and what pages (in our database) link to page i. We are thus able to construct a *web graph* with web pages represented as nodes and hyperlinks represented as directional edges.[19]

The seed set

The construction of the connectivity database first involved the identification of an initial sample or 'seed set' of web pages. The seed set for the present study was the top-60 ranked pages from two separate Google searches, the first using the phrase 'migration to Australia' and the second, two separate keywords 'migration' and 'Australia'.[20] We categorised each page in the seed set using the following organisational types: government agency; migration agent; embassy, consulate or high commission; industry association; personal home page; education or research facility; and other commercial organisations. We excluded pages that related to education or research about migration, as they are not of direct interest to potential migrants. We also excluded duplicate pages appearing in both lists. These exclusions left 32 pages from the 'migration to Australia' search and 40 pages from the 'migration and Australia' search.

We then combined these two lists, using a rank ordering approach. We started 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/ (No.1 on 'migration to Australia', and No. 3 on 'migration and Australia'), and www.migrationint.com.au/ (No. 6 on 'migration to Australia' and No. 1 on 'migration and Australia'). Hence, www.immi.gov.au/ was ranked number one in our seed set, and 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.

The rings

In the construction of our information space we cannot assume that potential migrants will only look at the pages returned by Google; it is highly likely that they will follow hyperlinks to other pages (and often, to other organisations). To account for this, our connectivity database contains two additional sets of pages that a potential migrant could encounter by following links from the seed set. We sent a web robot or crawler[21] into each of the pages in the seed set and used the results to generate two further 'rings' of pages. The '1st ring' is the set of pages returned from the web robot (i.e. the pages that pages in the seed set connect to), with each page satisfying two criteria: (1) each page in the 1st ring is not also represented in the seed set (i.e. it must be a 'new' page), (2) each page *i* in the 1st ring must be 'non-intrinsic' to (that is, not share the same domain name as) the page in the seed set that links to page *i*. The '2nd ring' set is then constructed in an analogous manner.[22]

Pages and page groups

The connectivity database contains 11,906 observations, with each observation representing a unique web page. However, we want to conduct analysis at the level of the organisation or functional grouping managing the site as a whole rather than at the level of the individual web page. We therefore aggregated web pages that come from the same organisation or functional grouping within an organisation into page groups (or 'sites'). In most cases, all pages with the same domain were placed into the same page group. For example, four pages from the website of the 'Victoria online' website of the Victoria government were aggregated into a single page group, www.vic.gov.au. However, in some cases we did distinguish between different functional groups within the same organisation. For example, our database contains

159 pages from the Australian Federal Government Department of Immigration and Multicultural and Indigenous Affairs (DIMIA) website and we aggregated these into 13 separate page groups including www.immi.gov.au/migration/family and www.immi.gov.au/migration/skilled. The reason we decided against placing all DIMIA pages into a single page group is that our analysis is attempting to identify differences in the web-based information available to migrants arriving under different schemes. It therefore makes sense to treat the DIMIA pages that pertain to these different schemes separately as far as data collection and analysis are concerned.[23]

Structure of the connectivity database

The connectivity database contains 7,755 page groups: 50 in the seed set, 1,142 in the 1st ring, and 6,563 in the 2nd ring. Using page groups as the unit of analysis, rather than pages, results in a 35 per cent decrease in the number of observations.[24] Even though the connectivity database was constructed with only two iterations of the web crawler, the depth of the path of outbound hyperlinks from a given page can be greater than two. For example, as shown in Figure 1, there is a path from DIMIA (www.immi.gov.au) to Amnesty International in Germany (www.amnesty.de) via the Migration Agents Registration Authority (www.themara.com.au), the Migration Institute of Australia (www.mia.org.au), and Amnesty International Australia (www.amnesty.org.au).

[Figure 1 about here.]

Figure 2 presents a screenshot of a cybermap with the DIMIA website as the root node (or "head" of the graph).[25] 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. There are

7,388 nodes in the DIMIA cybermap indicating that this site is connected (either directly or indirectly) to just about every other page group in the database.

[Figure 2 about here.]

WHAT INFORMATION DO POTENTIAL MIGRANTS FIND ON THE WWW?

Conceptualising the constructed information space as a core (the seed set) surrounded by two rings is useful because the ring structure represents the visibility of online information for potential 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). The pages in the 1st ring are going to be less accessible than those in the seed set (but still quite visible since each page in the 1st ring is, by definition, at most one step or degree of separation from at least one page in the seed set) while the pages in the 2nd ring will be even less visible.

In this section, we analyse the visibility of online information encountered by potential migrants by presenting a compositional analysis of the seed set and rings. We attempt 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 targeted on the WWW.

Composition of the seed set

In investigating what potential migrants may encounter on the web via the first level of searching, we can present information in three ways. The first is simply a list of the top ten websites by their organisation type (see Table 1).

The number one ranked site is the DIMIA site. It is the only government website included in the top 10 websites. Eight of the other sites belong to migration agents, that is, commercial businesses set up to assist people to migrate to Australia. Of these eight, four specifically target skilled migration. The other four list other types of migration such as working holiday visas, student visas, and family and spouse visas. The remaining website in the top 10 is the website of an Australian Embassy (in Austria).

[Table 1 about here.]

If we look at the 50 sites in the seed set by organisation type, it is evident that sites encountered by prospective migrants are most likely to be migration agents (48 per cent), followed by other commercial businesses (14 per cent) (see Table 2). Very few sites to make it into the top 50 are constructed by individuals, reflecting that very few non-mainstream players are visible. There are three sites which are personal homepages. A further three sites are defined as migration industry associations (which includes for example, professional association for migration agents).

[Table 2 about here.]

The TLD codes provide information on the type of organisation posting the information, and where the information is being provided. As found in the examination of the top 10 sites, most of the websites in the seed set are provided by commercial interests. Seventy-two per cent of the websites in the seed set are 'dot com'. This indicates the predominance of websites providing migration services. In this categorisation of data, personal websites tend not to have a generic TLD and are listed as 'unknown', but there is only a small per cent of these (six per cent). These results again indicate that the information which migrants encounter is largely from business rather than from individuals or community-based organisations.

[Table 3 about here.]

Perhaps surprisingly, less than half of the pages can be linked to an Australian information provider. That is, only 48 per cent of web pages in the seed set are '.au'. A large percentage of the web pages do not have a country TLD – this is not surprising given that many hostnames contain the generic TLD but no country code (that is, the page ends in .com or .org, for example). Also, some countries do not list a country code, the United States being the most notable, and in the database sites from such countries are coded as unknown. However, we can infer from the above that many of the web pages in the seed set are not provided in Australia. In the case of migration agents, many are international companies that provide migration services to potential migrants to many countries, not just Australia.

This investigation of the seed set shows that individual web pages are not highly visible when people search about information on migrating to Australia. The most prominent information is that supplied by migration agents who often specialise in business migration and who may not be based in Australia. Thus, the information does not appear to capture what Australia is like as a place to migrate to, and is more about the 'nuts and bolts' of applying for visas, and the services available from migration agents.

Composition of the rings sets

We did not categorise sites in the 1st and 2nd rings by organisational type as this would have involved looking at nearly 8,000 websites, which was not feasible in the context of this preliminary analysis.[26] Instead, we rely on the automatically-collected TLD information. This shows that the composition of the 1st and 2nd ring sets is different from the seed set. Commercial websites are not as prominent, although they still make up over half of sites in both ring sets (see Table 4). Government websites are much more prominent – this is because many seed sites (including government and migration agencies) point to government websites for information on migration issues. Sites are less likely to be identifiably Australian, which is perhaps an indication that the sites in the 1st and 2nd ring set are less relevant to prospective migrants to Australia than are those in the seed set.

[Table 4 about here.]

DISCUSSION AND CONCLUSION

Our quantitative characterisation of the online information environment encountered by prospective migrants to Australia allows us to conclude that Australia's online presence is largely defined by sites run by commercial entities and government agencies. Our preliminary work suggests that the information environment encountered by potential migrants appears to be heavily skewed towards skilled and business migrants and, further, that 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 consideration. We believe that this phenomenon results from a perception, held by the online information providers, 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'[27] 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 migrants 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 might otherwise go to Canada or the US. State government departments are also 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 try to win lucrative business associated with helping skilled and business migrants in the process of relocating to Australia.

In summary, our analysis suggests that the WWW is being used by commercial and government organisations to compete for skilled migrants. We also find that the “winner take all” phenomenon that has been observed in the context of politics on the web[28] 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.

For *providers of information,* we suggest that organisations must be aware of the processes in which they operate. An organisation operating in this information environment, say 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 shortcomings – a prospective migrant will not visit a website if they don’t know that it exists. Producers of information aimed at prospective migrants need to be aware of the implications of web topology (links) for the accessibility and visibility of online information.

For *consumers of information,* is the information they want actually available? 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. This is because Australia’s online presence (as perceived by prospective migrants) is slanted towards business people or highly skilled individuals.

Finally, for *those concerned with equality of access to online information and the digital divide* these findings provide an alternative way of analysing web use. 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. For example, men are more likely to use the web than women, and higher-income groups have been shown to have greater web usage rates than the poor. The focus on user surveys has also led to recent claims such as: ‘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...’.[29]

While web usage surveys can give important insights into the existence or otherwise of a digital divide, it is impossible to study the phenomenon adequately 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 can also 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. For them, it is appropriate 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.

ACKNOWLEDGEMENTS

This research has been supported by Australian Research Council Discovery Project DP0452051.

Table 1: The top 10 pages in the seed set[a] 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

[a] 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,[a] July 2004

Organisation type	Frequency	Per cent
Government Department	4	8.0
Australian Embassy, Consulate or High Commission	5	10.0
Migration Agent	24	48.0
Other commercial [b]	7	14.0
Personal homepage	3	6.0
Migration Industry Association	3	6.0
Other	4	8.0
Total	50	100.0

[a] Compiled from returns to two Google searches using first, the phrase ‘migration to Australia’ and second, the two separate keywords ‘migration’ ‘Australia’.

[b] ‘Other commercial’ includes Migration lawyers, homeloan services, and other business services.

Table 3: Composition of seed set by generic and country TLD, per cent

Generic	
com	72.0
org	10.0
gov	8.0
net	4.0

Unknown	6.0
Total	100.0
Country	
Australia	48.0
Other	8.0
Unknown	44.0
Total	100.0
Total N	50

Table 4: Composition of first and second ring sets by generic and country TLD, per cent

Generic	First ring	Second ring
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	100.0	100.0
Country	First ring	Second ring
Australia	42.6	38.4
United State	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	100.0	100.0
Total N	1,142	6,563

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

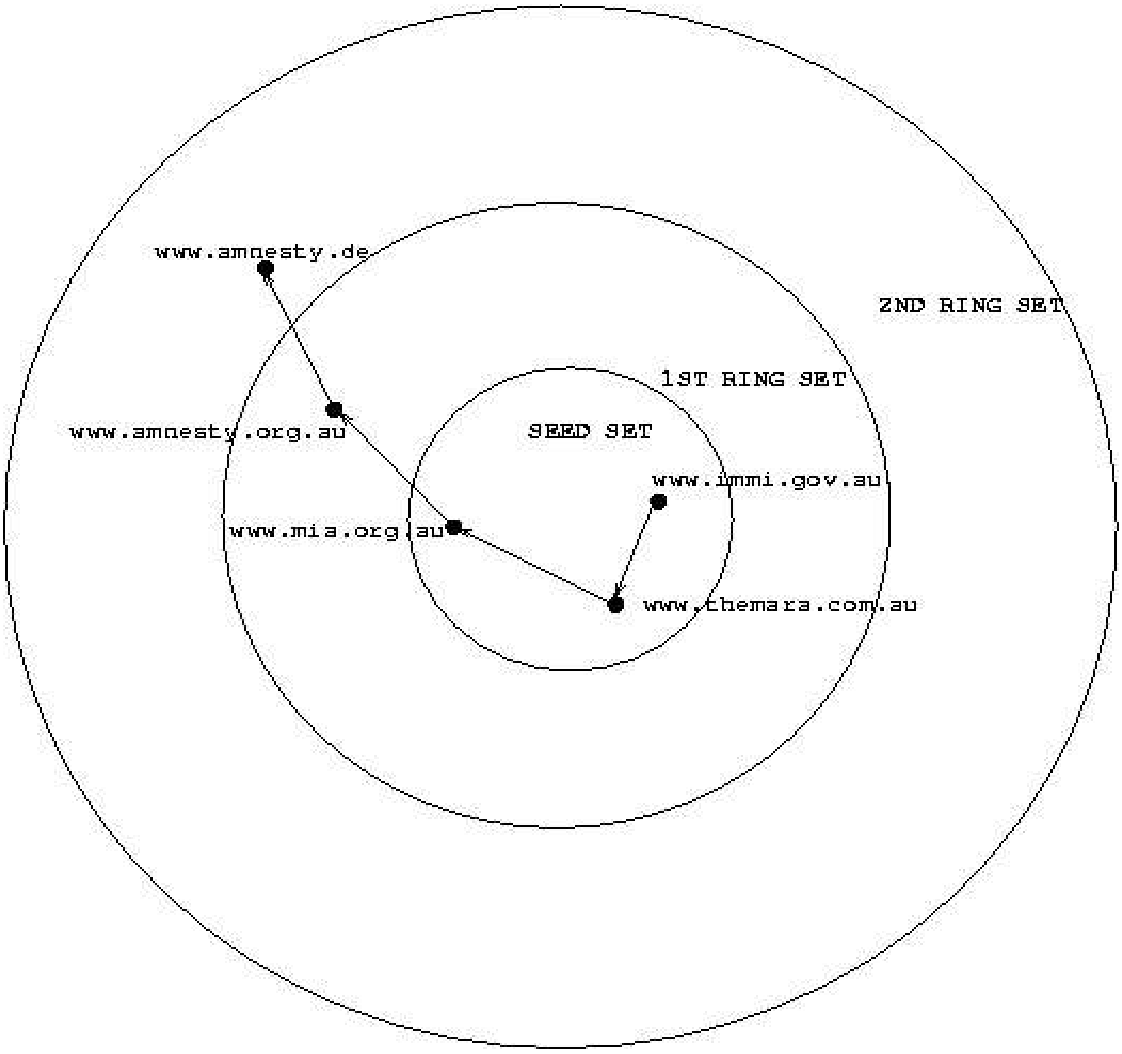
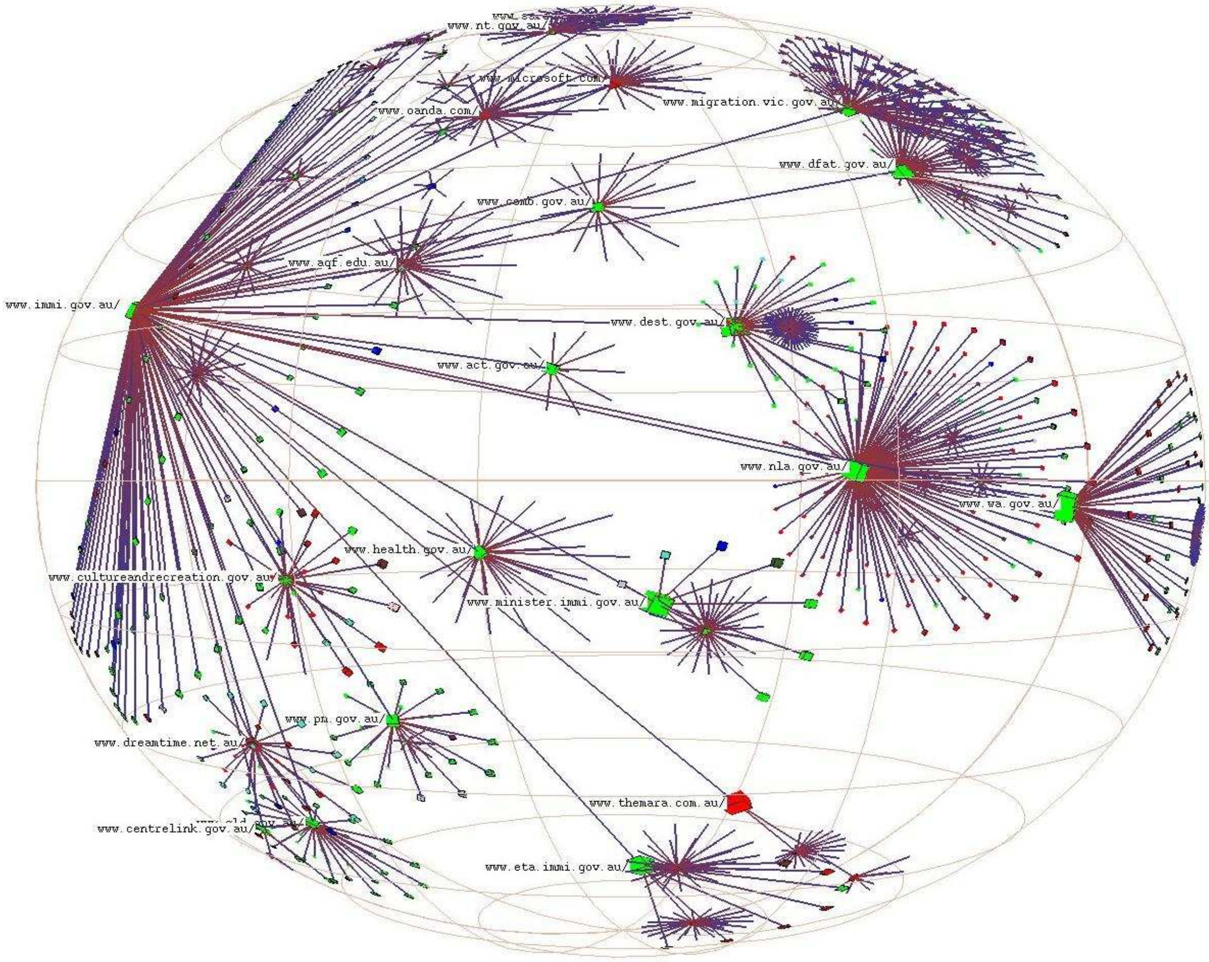


Figure 2: Connections between DIMIA and other websites in the database (H3Viewer)



REFERENCES

[1] M. Hindman, K. Tsioutsiouliklis, and J. Johnson ,' "Googlearchy": How a few heavily-linked Sites dominate politics on the web', mimeograph, Princeton University, 2003; R. Ackland and R. Gibson, 'Mapping far-right political party networks on the WWW', paper presented at the Political Science Program Seminar Series, Research School of Social Sciences, The Australian National University, 4th August 2004

[2] ibid.

[3] 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.

[4] Department of Immigration and Multicultural and Indigenous Affairs (DIMIA), 'Immigration update 2002–2003', Canberra, Research and Statistics Division, DIMIA, 2004.

[5] The remaining 17,597 settlers are termed 'Non-Program arrivals', and primarily consists of New Zealand citizens.

[6] ibid., p. 43

[7] P. McDonald and R. Kippen, 'Labor supply prospects in 16 developed countries, 2000-2050', *Population and Development Review*, vol. 27, no. 1, 2001, pp. 1–32

[8] L. Staehel, V. Ledwith, M. Ormond, K. Reed, A. Sumpter, and D. Trudeau, 'Immigration, the internet, and spaces of politics', *Political Geography*, vol. 21, 2002, pp. 989–1012

[9] H. Huijser, 'Internet hate: Exploring the limits of free speech', *Australian Mosaic*, vol. 5, 2004, pp. 27–28

[10] Hindman, et al., op. cit.

[11] ibid.

[12] There is an emerging industry in methods and practices aimed at maximising search engine rankings. *Adversarial information retrieval* has now been identified as a specific field of web research and the first international workshop on this topic will be held at the 14th International World Wide Web Conference (http://www2005.org/). Organisations with a profit motive are more likely to use these methods than are individuals maintaining their own websites.

[13] Staehel, et al., op. cit.

[14] C. Sunstein, *Republic.com*, Princeton, Princeton University Press, 2001

[15] Meta data is literally ‘data about data’.

[16] M. Dodge and R. Kitchin, *Mapping Cyberspace,* London, Routledge, 2001

[17] R. Ackland, ‘UberLink: software for analysing networks on the WWW (user guide)’, mimeograph, Canberra, The Australian National University, 2004

[18] 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). 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.

[19] 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.

[20] Note that we only collected web pages listed in the ‘organic’ (non-paid for) listing on Google.

[21] According to The Web Robots Pages (http://www.robotstxt.org/wc/faq.html, authored by M. Koster, accessed 28/10/2005), 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’.

[22] As noted in [R. Ackland, ‘Estimating the Size of Political Web Graphs,’ mimeograph, Canberra, The Australian National University, 2004] the construction of the connectivity database is analogous to constructing a snowball sample [see, for example, O. Frank and T. Snijders, ‘Estimating the size of hidden populations using snowball sampling’, *Journal of Official Statistics*, vol. 10, no. 1, pp 53-67]. A

few further points are worth noting. First, criterion 1 ensures that each page in the database is unique. Second, the web crawler will follow intrinsic links (links that are internal to the same website or domain) on a page in the seed set (up to a pre-specified number of links) but, as 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 1st ring set may in fact represent more than a single 'step' or 'jump' for a person following hyperlinks because the person may first have to 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 1st ring set.

[23] There may be other organisations in the database for which it would make sense to have multiple page groups (e.g. a migration agent may have a section of its website devoted to family migrants, while another section devoted to business migrants). For this preliminary analysis, we have just focused on creating multiple page groups for the DIMIA site, since this is the most important site in our database.

[24] Note 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 both the seed set and the 1st ring will be allocated to the seed set.

[25] The visualisation of hyperlinks is provided using the H3Viewer layout and graphical libraries described in T. Munzner, 'H3: Laying out large directed graphs in 3D hyperbolic space,' Proceedings of the 1997 Symposium on Information Visualization, October 20-21, 1997. Phoenix, AZ. See also http://graphics.stanford.edu/~munzner/. The visualisation uses 3D hyperbolic space, which exhibits the felicitous property of having more 'room' than standard 3D Euclidean space (and hence is useful for the display of large networks).

[26] Note that, for future research, we are investigating the use of sampling methods and also automatic content analysis tools as a possible means of finding more information about the sites in the 1st and 2nd rings of our dataset.

[27] D. Hoffman and T. Novak, ‘The evolution of the digital divide: How gaps in internet access may affect electronic commerce’, *Journal of Computer-Mediated Communication*, vol. 5, no. 3, 2000, available at: http://jcmc.indiana.edu/vol5/issue3/

[28] Hindman, et al., op. cit.

[29] J. Cole, ‘On superhighway to tomorrow, today’, *The Australian Higher Education Section*, 8 September 2004, p. 44