A final version of this is published as

Halavais, A. (2000) National borders on the world wide web. *New Media & Society 1*(3): 7-28.

National borders on the world wide web

Alexander Halavais

Abstract

The internet is often seen as a significant contributor to the globalization of culture and the economy. It is also seen as an inherently international medium, unimpeded by national borders and removed from the jurisdiction of the nation-state. This paper argues that although geographic borders may be removed from cyberspace, the social structures found in the 'real' world are inscribed in online networks. By surveying 4000 web sites, it is determined that the organization of the world wide web conforms to some degree to traditional national borders. Web sites are, in most cases more likely to link to another site hosted in the same country than to cross national borders. When they do cross national borders, they are more likely to lead to pages hosted in the United States than to pages anywhere else in the world.

Any recent discussion of globalization – economic or otherwise – is likely to make mention of the internet. The exponential growth of the internet in the United States, now being overtaken by growth in the rest of the world, has led many to question the relationship between a new global network and the future of sovereignty for the nation-state. McLuhan's 'global village' is frequently evoked and politicians strive to be seen as deliverers of information highways and wired schools. At the same time, many worry that instantaneous and ubiquitous transmission of cultural material may lead to a homogeneous world culture at best, and an American world culture at worst. That the internet leads inexorably toward globalization is often a foregone conclusion.

When evidence of internet-based globalization is presented it is all too often anecdotal. In the pages that follow, I will argue that national borders have a measurable effect on the topography of the world wide web. Surveying a sample of web pages allows us to determine the geographic distribution of hyperlinks. I will argue that these data do not support claims that 'cyberspace' exists as an anarchic unvariegated universe, unimpeded by national borders. Rather, while national borders seem to be less intrusive on the web than they are in earlier networked media, the resilience of cultural structures is demonstrated in the organization of this new medium.

National Borders and Communication

The question of identifying national borders on the internet is complicated by the fact that there is no clear agreement as to what 'national borders' are. By introducing a provisional definition of what national borders entail, we may be better able to detect homologies between these borders and the topography of the internet. A traditional view of the national border suggests itself as a starting point. In this conception, a national border is an imaginary boundary tied strictly to geographical territory in which a state's sovereignty may be exercised (Goodwin, 1974: 100). While such a definition is well suited to discussions between national governments, determining the sovereignty of nations by their territorial borders does not account for two vital ingredients of a nation: its people and their culture. People have always made connections across national borders, but improvements in communication and transportation technologies have made such connections far easier as the 20th century comes to a close. A political or legal sense of national borders ignores, by and large, these vital connections.

A second view of national borders is that they represent a gap in the totality of relationships between individuals.[1] Such an approach focuses more heavily on networks of association rather than more static or institutional notions of the state. John Burton (1972), for example, suggests that international political theory must change to recognize increases in transnational practices, especially among non-governmental groups. He argues that the traditional 'billiard-ball' model of international relations, in which countries of differing sizes act and react only in terms of a cohesive foreign policy, is lacking. He proposes, instead, the 'cobweb' model of international relations, in which layers of interaction define a world of highly complex interdependency. Such a view holds that while physical terrain may play a (decreasing) role in determining nationality, it is the imagined community, determined more by the propinquity of ideas than by the exigencies of physical distance, which defines nations and their borders.

Such a network or systems approach is very much in vogue among global communications scholars, due in no small part to the recent increased potential for distanced communication brought about by networking technology in general and the internet in particular. The idea that communications media have a substantial effect on the emergence of large-scale political formations is most often associated with Harold Innis (1972). However, it is Karl Deutsch's work in drawing relationships between measured communication flows and national boundaries that lays the foundation for the study presented here. Deutsch argues that although a number of factors contribute to establishing a nation, many of these are clearly measurable in the form of patterns of communication. In his words, we are able to draw from 'the observable ability of certain groups of men and women to share with each other a wide range of whatever might be in their minds, and their observable inability to share these things nearly as widely with outsiders' (Deutsch, 1953: 65). By examining the degree to which countries communicate internally and externally, and the character of that communication, we should gain some understanding of who becomes the 'outsider'; in other words, which 'peoples are marked off from each other by communicative barriers, by "marked gaps" in the efficiency of communication' (1953: 74). Measuring communicative flows on a large scale should provide some clues as to where these borders are being drawn in the collective imagination (Deutsch, 1956; also Janelle, 1991).

As one Singaporean minister has suggested, the modern nation is like a cell in a larger organism: porous in some respects, walled off in others, part of larger structures, containing sub-structures, ultimately in control of its own actions (Yeo, 1995: 23). By studying communication flows, we may gain some

understanding of the structure of imagined nations and some idea of where their borders lie. These borders may not be as arbitrarily exacting as those found on a world map, but by measuring the relationships between individuals and groups, we may arrive at a more dynamic and realistic measure of nations and their borders.

Measuring Borders in Cyberspace

It seems then, that determining the influence of national borders on the internet would be a fairly simple task. At least a gross measurement can be made by recording the flow of information on the internet across national frontiers. This process, however, is complicated by two factors. The first is the question of where, exactly, the internet ends. Certainly, we could say that the internet consists of the sum of all machines using the Internet Protocol, a set of standards that allows computers of different types to communicate. This, however, excludes systems that might make use of email that is then passed through a gateway to systems using the Internet Protocol. It also includes systems that use Internet Protocol but are isolated from the larger worldwide networks (December, 1996).

A second problem with measuring the flow of information on the internet is that it is fundamentally a distributed network. While much of the information on the internet flows through one or another 'backbone' it is far from easy to determine how this relates to the total amount of flow, or how this flow is related to users. Since the largest portion of bitflow is related less to content than to routing information (DNS lookups), a measure of flow alone would not necessarily reflect the amount of flow related directly to content exchange. While the Cooperative Association of Internet Data Analysis (CAIDA) is currently tackling problems of dataflow management for the major backbones, and there are opportunities for gathering similar information through server and network logging (Abrams and Williams, 1998), correlating these data with geographic information is not yet a viable option.[2]

Rather than attempting to measure all data flow on the net, it would be sensible to divide the problem into different ways in which the internet is used, or 'media classes', as December (1996) terms them. Arguments could be made for studying email, among the first and most widely used of internet technologies. Likewise, recent investigations of Usenet newsgroups have yielded very exciting depictions of how the medium is used and how groups organize themselves (Smith, 1999; Wittaker et al., 1998), and particularly how newsgroups might relate to national identity (Mackay and Powell, 1998). However, the world wide web is an appealing target for a number of reasons. First among these is the tremendous success of the web. It has been the driving force behind the popularization of the internet and is the fastest growing source of internet traffic (OECD, 1998). Part of this is due to the migration of other uses of the internet to the hypertext format. Gopher was subsumed by the web, and FTP, discussion lists, chat groups, and even email are increasingly indistinguishable from the web. As the web becomes less static and able to host more complex exchanges, it is likely it will become the predominant way of organizing information on the internet.

The web is an inherently public way of distributing content. Although there are increasingly areas of the web that require identification or payment to gain access (the New York Times and Wall Street Journal sites, for example), for the most part web pages are more like a billboard or graffiti, allowing messages to be easily placed in the public eye. Of course, email and newsgroups can also foster public communication, but the combination of collective construction and at least some stability makes the web

an especially attractive object of study. There is also something about the web's eponymous claim to 'world wide' distribution that begs the question of the degree to which it achieves this.

But the most intriguing aspect of studying the web is the structure it takes. Unlike many media, the web is constructed collectively and there exists no central authority to determine its overall structure. Though the creators of web pages are only a small fraction of the total number of users of the web, the barriers to creating material are relatively low and those creators can, for the most part, draw links to whatever other parts of the web they choose. While there have been a number of studies that have characterized various aspects of the content of the Web since its inception (Pitkow, 1998), structural descriptions of the medium have come only more recently. Those structural descriptions have tended to focus on the organization of the data in order to optimize navigation rather than as a collective expression through self-organization (Chakrabarti et al., 1998; Chelnokov and Zephyrova, 1997; Chen, 1997; Pirolli et al., 1996). Explorations of web structure, like that of Fagrell and Sørensen (1997) tend not to stress spatial relationships. Those that have investigated cyberspace in terms of a new cultural geography come closer to the aim taken here, but despite a strong interest in the effect of the web on national borders, little has been written about this aspect of 'cybergeography' (Curry, 1996; Dodge, 1998).

Given the networked organization of the world wide web, the natural approach would be to use the tools of network analysis to determine whether and where Deutsch's 'gaps' appear. Work in social network analysis has moved neatly to the internet because of the ready availability, in many cases, of transactional data (Garton et al., 1997). Email communication among members of an office staff, for example, can be recorded with minimal effort. Likewise, the web, at least on a small scale, is well-suited to network analysis. Network analysis would allow us to investigate the structure of the web in a more exploratory way, without an a priori hypothesis as to where borders exist.

While a network analysis of web pages would be ideal, in order to perform such an analysis a choice must be made between taking a sample that would have very little claim of generalizability and taking a census (or nearly so) of the web. A network analysis requires a network – that is, some connectedness among nodes – and any random or pseudo-random sample of the web of reasonable size is unlikely to define much of a network, if any. In order to collect a sub-network, a sample must be 'snowballed', choosing pages that are linked to a set of seed pages, and increasing the risk that the sample is not representative. On a small scale this is achievable but, particularly in attempting to determine the global structure of the web, this is not a very useful approach. A second approach, measuring the linkage structure of a large portion of the entire web is also a possibility.

Unfortunately, the resources required for this approach, both in terms of collection times and impact on the entire system, are extreme.[3]

For our purposes – questioning whether a specific set of borders exists or does not – a network analysis is also unnecessary. By taking a sample of web pages and determining where their hyperlinks lead, we should discover a set of patterns that indicate whether national borders have an effect on this distribution. This approach follows in some ways earlier attempts to measure international flows of information. Merritt and Clark (1977), for example, examine the international postal flows between 1890 to 1920 and show the development of gaps that predicted antipathies during World War I.

Of course, one could argue that the epistemic relationship between hyperlinks and data flow is tenuous. Naturally, it would be preferable to use bitflow data, were such data easily available on a large scale. Moreover, since all bits are not equal, it would be helpful to be able to characterize the content of that flow. Certainly, there is a need for more content-oriented qualitative analysis. Likewise, we could continue to rely on traditional user surveys to determine how often the internet is used to communicate internationally. But as Deutsch noted, 'Transactional research remains of crucial importance for the analysis of international relations. Interviews and survey data mainly tell us what people say; transaction data tell us what large numbers of them do' (Deutsch, 1979: 153).

Given the difficulty in drawing out this flow data, tracing hyperlinks is a reasonable and necessary first step in charting the structure of the web. It is tempting here to interject a McLuhanism: 'the user is the content' (in Levinson, 1999: 39). There are others who agree. There is some indication that Douglas Englebart, often attributed with presenting one of the earliest visions (and models) of hypermedia, saw hypertext structures as social structures (Bardini, 1997; Englebart, 1997), as have more recent researchers (Jackson, 1997; Smith, 1999).

We might think about hyperlinks as being analogous to roads, telephone lines or citations. Of course, roads alone would not tell you all you might want to know about the flow of people – some roads are used more than others. But where these roads were established demonstrates a social need of some sort and a road map certainly provides us with some indication of social geography. Likewise, telephone infrastructure can help identify social patterns, as Herbert Casson suggested in the early part of this century. He predicted that Bell Telephone's 'foresight department', then tasked with analyzing telephone infrastructure development, may one day:

> . . . become the first real corps of practical sociologists, which will substitute facts for the present hotch-potch of theories. It will prepare a 'fundamental plan' of the whole United States, showing the centre of each industry and the main runways of traffic. It will act upon the basic fact that *wherever there is interdependence, there is bound to be telephony*; and it will therefore prepare maps of interdependence, showing the widely scattered groups of industry and finance, and the lines that weave them into a pattern of national cooperation. (Casson, 1910: 96–7, original italics)

Finally, we might draw a parallel to citation analysis. While you may not follow every citation listed in the bibliography that follows this article, it would be fair to argue that those citations draw a connection (Paisley, 1990). Scholarly articles describe a citation structure arrived at collectively by their authors. Even without knowing the context of the citations or the number of people following the citations, an analysis can provide some idea of the structure of a field. It is natural to bring the ideas of co-citation to the web, as Larson (1996) has done, and to treat them as evidence of some kind of agency within a collective.

Given the alternatives, measuring the linkage structure of the world wide web clearly seems to be the best way to begin investigating the relationship of national borders to the internet. By comparing linkage data from a number of sites, we should be able to infer the impact of geographic borders. While simply comparing linkage patterns to established national borders is a more modest project than an exploratory investigation of gaps within the web at large, it represents a solid first step toward more ambitious investigations.

Collecting a Sample

The process of assembling linkage data on a large sample of web pages, though theoretically simple, faces a number of hurdles. The first, and in many ways the most difficult problem facing the researcher interested in the web is obtaining a useful sample. Because the web is constructed without central controls, a truly random sample of web pages, or even a reasonable approximation thereof, is unobtainable within the foreseeable future. Approaches by others who have required large samples from the web have varied. For certain applications, a domain-specific sample may be obtained simply by searching on appropriate keywords through search engines (Larson, 1996). Given that this survey aimed at a global view of the web, such an approach would be inappropriate. Bharat and Broder (1998) obtained their sample by first assembling a sample of Yahoo! pages to determine the number of given words in web documents, then using those words to query different search engines. While certainly a reasonable approach for their objectives (that is, measuring the coverage of various engines, and estimating the size of the web from these), such a process would yield very little in the way of a better sample for the survey undertaken here. Finally, many attempts to determine the character of the 'average' web page are based on incomplete but extensive samples provided during the construction or operation of a search engine (Bray, 1996; Brin and Page, 1998; Woodruff et al., 1996). Because of the competitive nature of the search engine business, complete indices are considered proprietary and not normally available to the researcher. Moreover, assembling such data on an individual basis would unnecessarily tax the infrastructure (Cerf, 1991).

Table 1. Sample compared to the world-wide distribution of registered domains and hosts

COUNTRY	HOSTS[4]	% OF TOTAL	SITES IN SAMPLE	% OF TOTAL	LINKED TO IN SAMPLE	% OF TOTAL
United States	20,623,323	69.5	2,874	78.0	41,209	77.2
Germany	994,926	3.4	101	2.7	1,166	2.2
United Kingdom	987,733	3.3	157	4.3	1,586	3.0
Sweden	319,065	1.1	62	1.7	623	1.2
Australia	665,403	2.2	43	1.2	861	1.6
Netherlands	381,172	1.3	49	1.3	546	1.0
Japan	1,168,956	3.9	27	0.7	410	0.8
Canada	839,141	2.8	88	2.4	1,241	2.3
Switzerland	114,816	0.4	18	0.5	288	0.5
Brazil	117,200	0.4	9	0.2	123	0.2
Italy	243,250	0.8	37	1.0	357	0.7
New Zealand	169,264	0.6	7	0.2	63	0.1
South Africa	122,025	0.4	8	0.2	98	0.2
France	333,306	1.1	25	0.7	262	0.5
Norway	286,338	1.0	20	0.5	1	0.0
Other Countries	2,303,693	7.8	161	4.4	4,533	8.5
TOTAL	29,669,611		3,686		53,367	

The sample used in this survey consisted of 4000 sites drawn from Excite's Webcrawler search engine in the early part of 1998, using its web-based 'roulette' page which provides a sample of pages drawn randomly from the engine's index. This approach, while certainly flawed, has been suggested as a reasonable approach for obtaining a sample of convenience (Lock, 1997). The sample provided may be

skewed toward American culture and English-speaking web sites; and lacking anything approaching a census of the web, it is difficult to estimate how biased the sample is in this regard. In terms of the physical location of the web sites surveyed, the sample seems to be a good approximation of other measures of the international distribution of the web: for example, surveys of registered domains and web servers (see Table 1).

A specific definition of a 'site' was taken: only links that were proximate to the 'base URL' of the page indexed in the sample were included.[5] This represents a compromise between choosing an individual page as a unit of analysis and gathering all pages within a given domain. The former, a more popular choice for surveys of web content, ignores the intent of the author that the pages be bound together as a whole. It therefore assumes that all hyperlinks are external hyperlinks. Also, an analysis of top-level pages alone would likely contain far less links than were found by digging down into the hierarchy of each site. At the other extreme, domains were not selected as they often contain more than a single web site and these sites may or may not be directly interconnected. This is particularly true of personal home pages at a business or university which may not be connected to other sites in the same domain. As a result, the sample contains several hundred individual sites within large domains like Geocities, Tripod, AOL, and Angelfire.[6] By picking those pages that were archived by the search engine and working downward in the hierarchy of the site from this point, we reach a reasonable approximation of a representative web site.

Once URLs for the 4000 sites in the sample were assembled, a 'crawler' was created to search through the pages in each site and record relevant data. Search engines often use crawlers (also referred to as 'spiders' or 'bots') to automatically collect data from the web. Fagrell and Sørensen (1997) made use of a crawler very similar to the one employed here, though their approach was directed more toward determining the characteristics of the average web page than the destination of various hyperlinks. The crawler used here, created in the Python scripting language, visited each of the 4000 sites in the sample and recorded the first 50 pages at each location (transversing the links in a breadth-first fashion, accessing all of the links on a given page before moving on to links on subsidiary pages), resulting in information from a total of 45,457 pages being gathered. These pages were then parsed to determine the destination of all external hyperlinks; links to other pages within the site were ignored. These hyperlinks were then coded for location, relying on two-letter top level domains (TLD) to determine the country in which each page was registered. Thus a site with the URL of http://www.yahoo.ca would be coded as being a 'Canadian' site, based on the final two letters.[7] Those sites within the three-letter generic top level domains (also referred to as gTLDs; .com or .edu, for example) were checked against the WhoIs registry to determine the country of origin. All but 6 percent of these were registered to a US billing address, a reasonable approximation of their hosting location (Zook, 1998). Fully a quarter of the .com domains newly registered in 1997 were to a foreign address ('It's a Wired World', 1997), so it is reasonable to assume that the number of foreign sites using three-letter gTLDs will continue to increase. Once the destination of each hyperlink was determined, the total percentage of hyperlinks from the site to various countries was recorded.

The data from hyperlinks alone would suffice in providing evidence of concentration of domestic rather than international linkages. However, the data collected can also provide at least some indication of how subject matter and language affect those international linkages. To that end, the text for each site was coded for its broad category and for the language(s) used (see Tables 4 and 5). The categories listed were derived from the top level category labels used on the Yahoo! directory, with the addition of several

categories that merited special attention. For each category, a model text was assembled consisting of 30 pages selected from the Yahoo! directory.[8] Sites were categorized by comparing word frequencies between the gathered text and the model texts for each category. This allowed a cross-tabulation between the language used (or number of languages used), the general topic of a site, and the percentages of international hyperlinks for a given site.

Analysis

An analysis of the data provided by this survey leads to findings in two areas. First, while the world wide web is a very international medium, the number of hyperlinks that cross international borders are significantly less than those that link to sites within the home country. Second, links are far more likely to be directed toward the United States than toward any other country, though this appears to be due in large part to the imbalance in the number of sites hosted in each country.

If the destination of hyperlinks is aggregated for the 12 countries best represented by the sample, it becomes clear that domestic links are far more common than international links (Figure 1). With the exception of those hosted in Canada, sites were more likely to link to another site within the same country than to cross national borders (Table 2). This is not at all what we would expect if the world wide web were, indeed, an undifferentiated network. Rather, we would expect a fairly even distribution of linkages across the web. The tendency to link to domestic sites is particularly significant, given that outside of domain names in the URL there is very little to indicate or to restrict the user from crossing national boundaries while surfing the web.

Of course, web sites are not evenly distributed among countries, so we cannot assume that links will be. The initial diffusion of the web was certainly heaviest in the United States, and the relative maturity of the web in the US means that a majority of pages are hosted by American servers. Using the distribution of hosts as a guide, we would expect, for example, about 70 percent of all links on the web to lead to the US and about 1 percent of all links to lead to Japan. In fact, as shown in Table 3, the United States receives a lower percentage of links than we would expect from sites around the world (except from those sites located within the US). Table 3 (and Figure 2) show how the percentage of linkages between countries differs from the distribution of host machines. While the percentage of linkages to the United States is quite high in absolute terms, it is unremarkable when the distribution of the web is taken into account.

There remains a bias toward domestically produced content in the US, but this bias is fairly small when compared to the relatively inwardly linking web in France and Japan. We are left with an ambiguous picture. A majority of web content is created in the United States and this content is linked to nearly as frequently as material produced indigenously around the world.[9] However, a very large part of the bias toward the United States seems to be as a result of the distribution of content at this stage of the internet. If this is, as this survey shows, an accurate depiction of today's web, we might expect to see this disparity diminish as more of the world begins to use the internet. On the other hand, the population, economic power, and technological position of the United States makes it unlikely that countries like Canada or individual nations in Europe will be able to challenge the centrality of America on the web in the very near future, especially in terms of networking infrastructure (Evagora, 1997; OECD, 1998).

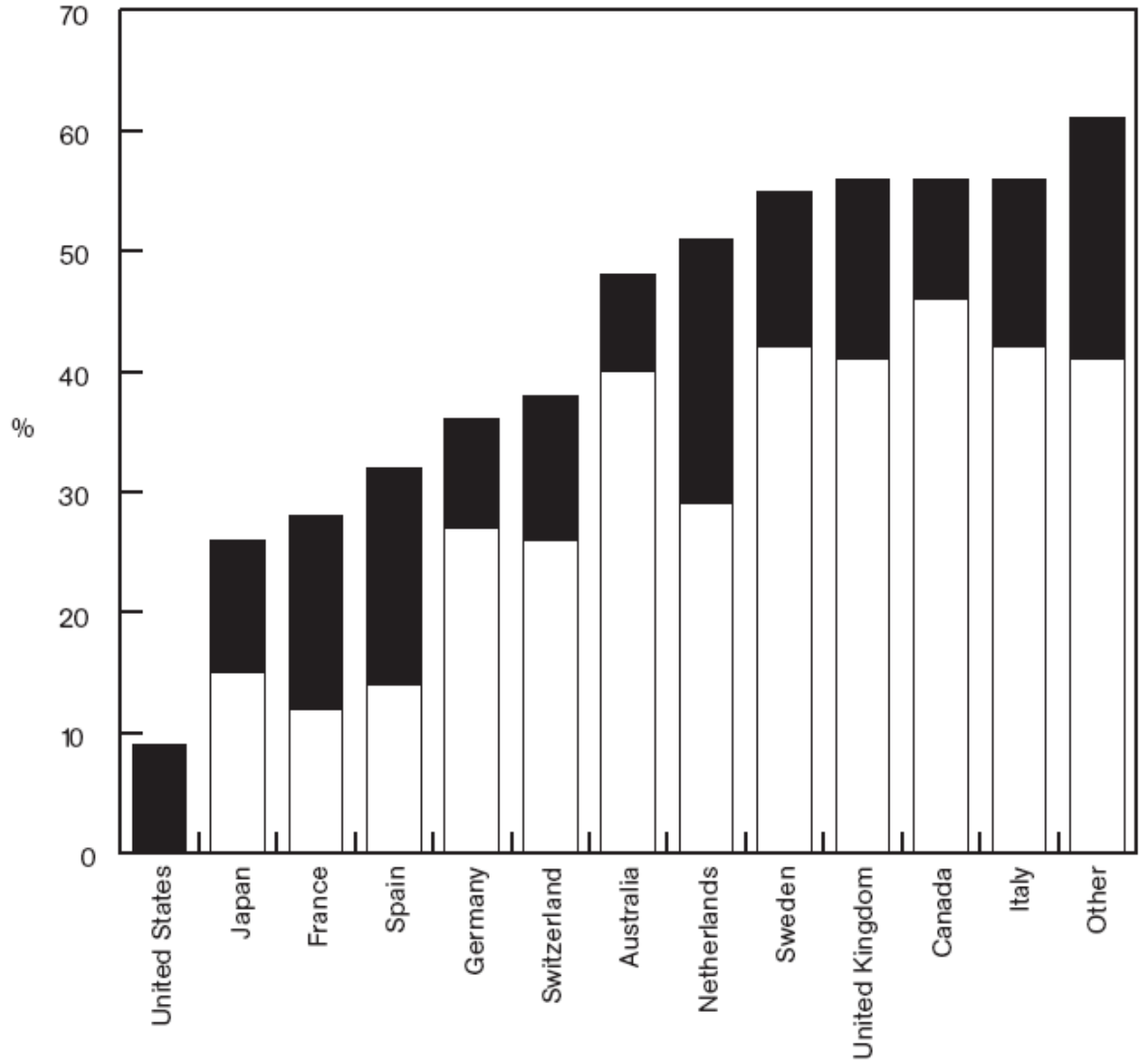

Figure 1: Percentage of international hyperlinks from web sites in 12 countries. ■ links to non-US sites; □ links to US sites.

The structure of content on the web may prove to be far more dynamic, especially if the rapid changes in demographics of web users are any indication of the volatility of this new medium (Bloomberg, 1998). As has already been noted by a number of technically as well as socially minded internet researchers, the measurement and mapping of the internet is of vital importance. In this vein, Tim Berners-Lee has called for 'parameters of measurement of restlessness and stability analogous to hormone levels or body temperature of the human organism' as an indication of how this structure is changing (Berners-Lee, 1997). Such measures should provide not only information about 'cyberspace' but about how networking affects real space.

Table 2: Distribution of links by country (12 largest countries in sample) (%)[10]

	Linking to (%)												
Sites in:	US	UK	CA	DE	AU	NL	SE	JP	IT	FR	ES	CH	OTHER
United States (us)	**90.7**	1.9	1.6	0.5	0.8	0.5	0.5	0.3	0.3	0.4	0.1	0.4	1.1
United Kingdom (uk)	42.6	**43.4**	1.1	1.3	1.0	1.3	2.6	0.1	0.5	0.6	0.5	1.0	3.1
Canada (ca)	48.2	2.4	**43.1**	0.2	1.7	0.1	0.5	1.5	0.0	0.2	0.0	0.1	1.9
Germany (de)	27.7	1.7	0.4	**63.5**	0.0	0.3	0.4	0.0	0.4	0.2	2.3	0.7	2.2
Australia (au)	39.7	0.6	0.5	0.4	**52.5**	0.5	0.5	0.2	0.5	0.1	0.0	0.3	3.6
Netherlands (nl)	29.7	7.3	1.0	4.3	0.7	**49.4**	0.4	0.1	0.1	0.1	0.0	0.3	5.9
Sweden (se)	43.3	2.8	1.0	1.8	1.0	0.7	**44.9**	0.4	0.1	0.2	0.0	0.6	2.5
Japan (jp)	15.1	0.3	0.4	0.0	0.9	1.0	0.0	**74.6**	0.0	0.1	0.0	3.1	0.9
Italy (it)	43.0	5.6	0.1	0.3	0.2	2.8	1.8	0.0	**42.7**	0.3	0.0	1.8	0.8
France (fr)	11.8	2.1	11.0	0.2	0.0	0.5	0.0	0.0	0.7	**71.9**	0.0	0.0	1.7
Spain (es)	14.3	1.2	0.1	0.0	0.1	0.3	0.0	0.0	0.4	10.0	**68.1**	0.3	1.2
Switzerland (ch)	26.1	6.2	0.0	2.5	0.0	0.0	0.0	0.0	0.0	1.7	0.0	**62.2**	1.3
Other countries[10]	40.9	1.7	0.8	1.8	1.8	1.1	0.8	0.3	0.2	0.3	0.0	1.4	**40.3**

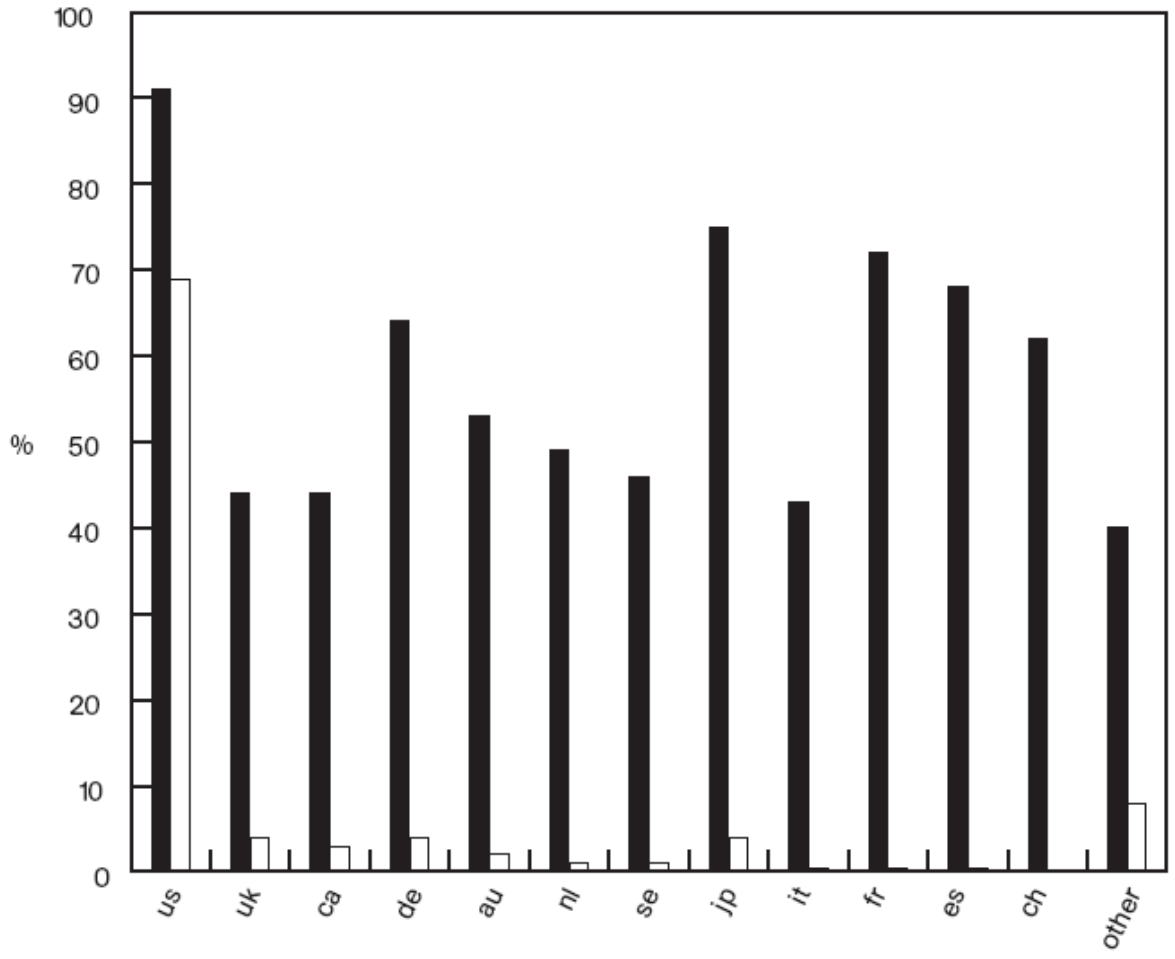

Figure 2: Domestic linkages in comparison to share of world's domains. ■ domestic hyperlinks; □ share of world's domains.

Strikingly clear in both Tables 2 and 3 is the tendency of sites to link domestically rather than internationally. When compared to two traditional networked media – the telephone and postal systems – the web appears to be much more internationalized. Yet the degree to which there are gaps at national borders is remarkable; even more so when it is noted that unlike the postal and telephone systems, the web provides no differential pricing for domestic and international linkages. As cost is reduced, we would expect the network to become increasingly interconnected. In the case of the web, however, it is clear that there are other, non-economic barriers to distanced networking. While a hyperlink from Paris to Nice may cost the same as one from Paris to Tokyo, the former is far more likely. The exact location of the most significant gaps in communication cannot be determined by the approach taken here, but it is clear that these gaps are to a degree correlated to national borders. When other borders are removed, social homophily guides the selection of necessarily scarce hyperlinks (Van Alstyne and Brynjolfsson, 1997).

Table 3: Difference between expected and actual percentages of hyperlinks (%)

	Linking to (%)												
Sites in:	US	UK	CA	DE	AU	NL	SE	JP	IT	FR	ES	CH	OTHER
United States (us)	**21.2**	−1.4	−1.2	−2.9	−1.4	−0.8	−0.6	−3.6	−0.5	−0.7	−0.5	0.0	−6.1
United Kingdom (uk)	−26.9	**40.1**	−1.7	−2.1	−1.2	0.0	1.5	−3.8	−0.3	−0.5	−0.1	0.6	−4.1
Canada (ca)	−21.3	−0.9	**40.3**	−3.2	−0.5	−1.2	−0.6	−2.4	−0.8	−0.9	−0.6	−0.3	−5.2
Germany (de)	−41.8	−1.6	−2.4	**60.1**	−2.2	−1.0	−0.7	−3.9	−0.4	−0.9	−1.7	0.3	−5.9
Australia (au)	−29.8	−2.7	−2.3	−3.0	**50.3**	−0.8	−0.6	−3.7	−0.3	−1.0	−0.6	−0.1	-3.6
Netherlands (nl)	−39.8	4.0	−1.8	0.9	−1.5	**48.1**	−0.7	−3.8	−0.7	−1.0	−0.6	−0.1	−1.3
Sweden (se)	−26.2	−0.5	−1.8	−1.6	−1.2	−0.6	**43.8**	−3.5	−0.7	−0.9	−0.6	0.2	−4.7
Japan (jp)	−54.4	−3.0	−2.4	−3.4	−1.3	−0.3	−1.1	**70.7**	−0.8	−1.0	−0.6	2.7	−6.3
Italy (it)	−26.5	2.3	−2.7	−3.1	−2.0	0.5	0.7	−3.9	**41.9**	−0.8	−0.6	1.4	−6.4
France (fr)	−57.7	−1.2	8.2	−3.2	−2.2	−0.8	−1.1	−3.9	−0.1	**70.8**	−0.6	−0.4	−5.5
Spain (es)	−55.2	−2.1	−2.7	−3.4	−2.1	−1.0	−1.1	−3.9	−0.4	8.9	**67.5**	−0.1	−6.0
Switzerland (ch)	−43.4	2.9	−2.8	−1.6	−2.2	−1.3	−1.1	−3.9	−0.8	0.6	−0.6	**61.8**	−5.9
Other countries	−28.6	−1.6	−1.1	−0.6	−0.4	−0.2	−0.3	−3.6	−0.6	−0.8	−0.6	1.0	**33.1**

As seen in Table 4, the degree to which these gaps are present differs depending on the subject matter found on the site. The most internationalized pages tend to be those related to the international scholarly community, while the least tend to be pages related to governmental bodies. News, sports, and (strangely enough) travel tend to be less oriented toward international hyperlinks than web sites centered on other topics.

We might approach this information in two ways. We might begin by considering the web as an indicator of the global social environment. For example, scientific and scholarly communities have long been international in nature, as have certain political movements. As the topical index suggests, these groups have quickly migrated to the new medium. Other groups have only recently seen an increase in the need for transnational communication. For example, the elimination of many economic impediments has driven even small businesses into the international market.

However, the web has also provided excess capacity for transnational practices. While some Americans may be 'bowling alone', many others are taking up hobbies and interests – from anime to macramé – for which they find support from outside of their physical communities. Many businesses approach the web as a cheap source of advertising or another venue for sales and 'stumble into' the international aspect of the medium.[11] As users come to depend on the web, they enter into negotiation with its conventions, adopting those they like and adapting to those they do not. Because of the reciprocal relationship between

public conceptions of the web and its actual structure, the future of that structure remains difficult to predict.

Table 4: Percentage of foreign links by topic area (includes only sites with external links)

Topic Area	Foreign Links (%)	Total Sites
Science and research	38	75
Internet and computers	34	95
Political	32	24
Recreation	28	122
Personal	27	310
Business	25	515
Education	24	142
Arts and entertainment	24	247
Social and religious groups	21	128
News	20	77
Sports	19	90
Travel	19	63
Health	19	46
Regional	17	48
Government	9	52

Table 5: Percentage of foreign links by language (includes only sites with external links)

Language	% of Total Sample	% Linking Internationally
Only English	92	23
Some English	95	24
Single non-English Language	5	32
Multiple Language	3	75

The preponderance of the sites in this sample contain pages in English, as shown in Table 5. The overwhelming presence of English on the web is a cause for concern, given the potential of language as perhaps the most powerful of ways to establish borders in this new medium (Castells, 1997: 52). As noted above, the sample may be slightly skewed toward sites in countries in which the English language is dominant, so some caution should be taken before inferring too much from the large number of English-language sites. However, the percentage of English-language sites that link domestically is instructive. We might expect English-language sites to be more likely to link internationally if English is indeed the new global lingua franca. One possible explanation for this is that commercial sites – which are far more likely to be 'dead-end' sites, without hyperlinks outside the site (Terveen and Hill, 1998) – appear more often in English. Less surprising, perhaps, is the fairly large amount of international linkage associated with multilingual sites.

Conclusions

The findings presented here lead to an immediate set of conclusions and a more forward looking set of suggestions. The first of these is related to a reading of the internet at a particular point in time. The novelty of the internet forces discourse about it to the extremes, and hyperbole abounds when questions of national borders, sovereignty and the internet are addressed. The survey undertaken here demonstrates

clearly that, as with earlier 'new media', this technology is both 'so new that people can't even imagine it' and 'never so new as people imagine' (Nord, 1986). On one hand, while the internet incorporates little in the way of technological, regulatory or economic impediments to transnational interconnections, the web demonstrates that national cultures continue to exert a substantial influence on how these connections are made. While national borders may be eroded, they certainly remain significant.

On the other hand, when compared to other media, the internet is considerably more internationalized. If you examine postal flows, none of the 12 countries considered here receives more than 5 percent of its total letters from abroad (UPU, 1997). The United States, the most insular (in absolute terms) of the countries considered here, has over 9 percent of its links fetching information from abroad, while other countries have much higher rates of international content. This presents a novel opportunity for people to be exposed to information and ideas from outside their own national cultures (to the extent that such can be said to exist). I would suspect that email and other more 'personal' uses of the internet would be more geographically localized and this presents an interesting area of inquiry.

The existence of national borders, though perhaps not in the more traditional sense, has important policy impacts. The chief argument against national regulation of the internet is that it is inherently global. This survey indicates that for the web, national borders are neither absent nor absolute. Certainly there are significant challenges to designing regulation for a social environment that is less reliant on geography (Lenk, 1997). However, ridiculing policy-makers who claim that the internet can be segmented or controlled ignores the crucial impact of social structure on the structure of the medium.

In her dissent to the Supreme Court's striking down of the Communications Decency Act, Justice Sandra Day O'Connor argued that cyberspace could indeed be segmented if the desire to do so was made clear in social policy (Reno v. ACLU, 1996).[12] In making such an argument, she explicitly drew on ideas presented by Lawrence Lessig, a law professor presently at Harvard University. Lessig (1996) argues that there are a number of elements that lead to structures in the networked environment. While some of these are technological, most of them emerge as social (and often commercial) constructs. Legal borders – national and otherwise – emerge as social conventions. As such, they need not rely expressly upon geography. As the internet becomes more socialized, law will develop that takes into account the new borders of cyberspace (Johnson and Post, 1997). The present position of the United States as central to the web may also be a cause for alarm. The future of this distribution is in no way certain – it may go the way of earlier mass media and the US may maintain a central position on the internet, as Herbert Schiller and others have argued (Gillespie and Robins, 1989; Schiller, 1992, 1995). However, the widely noted speed of diffusion of the internet outside of the United States, and the relatively open linking of US sites to sites abroad is likely to present a significant challenge to the centrality of the United States on the web (Maherzi, 1997: 46–7).

More important than these conclusions, which given the ephemeral nature of the internet remain necessarily of the moment, are the conceptual

underpinnings of this study. An attempt to describe the social impacts of the internet must include some indication of the structure of this medium. A description that suggests the internet is an undifferentiated space outside of 'real' space – as many popular and academic accounts do – must be approached with some skepticism. We must recognize that 'social borders have their own cartographies' and go about mapping these structures (Harvey, 1996: 282). This is in no way a new concern – Georg Simmel noted

that modernity has provided any number of examples of 'a group whose cohesion depended upon geographic and physiological factors, terminus a quo, [being] entirely replaced by a group whose cohesion was based on purpose, on factual considerations, or, if one will, on individual interests' (Simmel, 1955: 128). The internet provides a very promising way to observe how these borders and groups evolve. While the future of internetworking will most certainly surprise us, the need to investigate the social and informational structure of the medium will continue to remain among the most important tasks of the researcher.

References

Abrams, M. and S. Williams (1998) 'Complementing Surveying and Demographics with Automated Network Monitoring', WWW Journal, 3.

Arnum, E. and S. Conti (1998) 'Internet Deployment Worldwide: The New Superhighway Follows the Old Wires, Rails, and Roads', INET '98, URL (consulted April 1999): http://www.isoc.org/inet98/

Bardini, T. (1997) 'Bridging the Gulfs: From Hypertext to Cyberspace', Journal of Computer Mediated Communications 3(2), URL (consulted April 1999), http://www.ascusc.org/jcmc/

Berners-Lee, T (1997) 'World-wide Computer: The Next 50 Years: Our Hopes, Our Visions, Our Plans', Communications of the ACM 2(40): 57.

Bharat, K. and A. Broder (1998) 'Estimating the Relative Size and Overlap of Public Web Search Engines', WWW7 Proceedings, Sydney, Australia.

Bharat, K., A. Broder, M. Henziger, P. Kunar and S. Venkatesubramanian (1998) 'The Connectivity Server: Fast Access to Linkage Information on the Web'. Computer Networks and ISDN Systems 30(1–7): 469–77.

Bloomberg (1998) 'Internet Executives Say Fastest Growth is from Outside US', South China Morning Post, 12 May, Internet edn, URL (consulted July 1998): http://scmp.com

Bray, T. (1996) 'Measuring the Web', WWW5 Proceedings, Paris, France.

Brin, S. and L. Page (1998) 'The Anatomy of a Large-Scale Hypertextual Web Search Engine', Computer Networks and ISDN Systems 30: 107–17.

Burton, J. (1972) World Society. Cambridge: Cambridge University Press.

Casson, H. (1910) The History of the Telephone. Chicago: AC McClurg & Company.

Castells, M. (1997) The Power of Identity (The Information Age: Economy, Society and Culture, Volume II. Cambridge, MA: Blackwell.

Cerf, V. (1991) 'Guidelines for Internet Measurement Activities', Network Working Group Request for Comments no. 1262, URL: http://www.cis.ohio-state.edu/hypertext/ informatoion/rfc.html or http://www.faqs.org/rfcs.

Chakrabarti, S., B. Dom, P. Raghavan, S. Rajagopalan, D. Gibson and J. Kleinberg (1998) 'Automatic Resource Compilation by Analyzing Hyperlink Structure and Associated Text', WWW7 Proceedings, Sydney, Australia.

Chelnokov, V. and V. Zephyrova (1997) 'Collective Phenomena in Hypertext Networks', paper presented at Hypertext 97, Southampton, UK.

Chen, C. (1997) 'Structuring and Visualising the WWW by Generalised Similarity Analysis', paper presented at Hypertext 97, Southampton, UK.

Curry, M. (1996) 'Cyberspace and Cyberplaces: Rethinking the Identity of Individual and Place', paper presented at the International Association for Mass Communication Research Conference, 18–22 August, Sydney.

December, J. (1996) 'Units of Analysis for Internet Communication', Journal of Communication 46(1): 14–29.

Deutsch, K. (1953) Nationalism and Social Communication: An Inquiry into the Foundations of Nationality. Cambridge, MA: The Technology Press of MIT; and New York: John Wiley & Sons.

Deutsch, K. (1956) 'Shifts in the Balance of International Communication Flows', Public Opinion Quarterly 20 (Spring): 143–60.

Deutsch, K. (1979) Tides Among Nations. New York: The Free Press.

Dodge, M. (1998) 'The Geographies of Cyberspace', paper presented at the Association of American Geographers Conference, Boston, March.

Drake, W. (1993) 'Territoriality and Intangibility: Transborder Data Flows and National Sovereignty', in K. Nordenstreng and H. Schiller (eds) Beyond National Sovereignty: International Communication in the 1990s, pp. 259–313. Norwood, NJ: Ablex.

Englebart, D. (1997 [1963]) 'A Conceptual Framework for the Augmentation of Man's Intellect' in P. W. Howerton and D. C. Weeks (eds), Vistas in Information Handling Volume 1: The Augmentation of Man's Intellect by Machine, pp. 1–29. Washington, DC: Spartan Books.

Evagora, A. (1997) 'World Wide Weight', Tele.com (August), URL (consulted April 1999): http://www.teledotcom.com/

Fagrell, H. and C. Sørensen (1997) 'It's Life Jim, But Not As We Know It!', paper presented at IRIS 20 (Social Informatics), Hankø Fjordjotel, Norway, 9–12 August.

Farrell, C., M. Schulze, S. Pleitner and D. Baldoni (1994) 'DNS Encoding of Geographical Location', Network Working Group Request for Comments no. 1712, URL: http://www.cis.ohio-state.edu/hypertext/information/rfc.html or http://www.faqs.org/rfcs.

Garton, L., C. Haythornthwaite and B. Wellman (1997) 'Studying Online Social Networks', Journal of Computer Mediated Communication 3(1), URL (consulted April 1999): http://www.ascusc.org/jcmc/

Gillespie, A. and K. Robins (1989) 'Geographical Inequalities: The Spatial Bias of the New Communications Technologies', Journal of Communication 39(3): 7–18.

Goodwin, G. (1974) 'The Erosion of National Sovereignty?' in G. Ionescu (ed.) Between Sovereignty and Integration, pp. 100–17. New York: Wiley.

Harvey, D. (1996) Justice, Nature & the Geography of Distance. Malden, MA: Blackwell.

Innis, H. (1972) Empire and Communication. Oxford: Clarendon Press.

'It's a Wired World' (1997) PC Magazine Online, 26 September, URL (consulted June 1998): http://www.zdnet.com/pcmag/

Jackson, M. (1997) 'Assessing the Structure of Communication on the World Wide Web', Journal of Computer Mediated Communication 3(1), URL (consulted April 1999): http://www.ascusc.org/jcmc/

Janelle, D. (1991) 'Global Interdependence and Its Consequences', in S. D. Brunn and T.

R. Leinbach (eds) Collapsing Space and Time: Geographic Aspects of Communication and Information, pp. 49–81. London: HarperCollins Academic.

Johnson, D. and D. Post (1997) 'The Rise of Law on the Global Network', in B. Kahin and Charles Nesson (eds) Borders in Cyberspace: Information Policy and the Global Information Infrastructure, pp. 3–47. Cambridge, MA: MIT Press.

Larson, R. (1996) 'Bibliometrics of the World Wide Web: An Exploratory Analysis of the Intellectual Structure of Cyberspace', American Society for Information Science 1996 Proceedings, 19–24 October, URL (consulted April 1999): http://sherlock.berkeley.edu/asis96/asis96.html

Lenk, K. (1997) 'The Challenge of Cyberspatial Forms of Human Interaction to Territorial Governance and Policing', in B. Loader (ed.) The Governance of Cyberspace: Politics, Technology and Global Restructuring, pp. 126–35. London: Routledge.

Lessig, L. (1996) 'Reading the Constitution in Cyberspace', Emory Law Journal 45(869).

Levinson, P. (1999) Digital McLuhan: A Guide to the Information Millennium. London: Routledge.

Lock, R. (1997) 'Internet Resources for Teaching Statistics', paper presented for the 2nd World Conference of the International Association for Statistical Computing, Pasadena, California, February.

Mackay, H. and T. Powell (1998) 'Connecting Wales: The Internet and National Identity', in B. Loader (ed.) Cyberspace Divide: Equality, Agency and Policy in the Information Society, pp. 203–18. London: Routledge.

Maherzi, L. (1997) World Communication Report: The Media and the Challenge of the New Technologies. Paris: UNESCO.

Merritt, R. and C. Clark (1977) ‘An Example of Data Use: Mail Flows in the European Balance of Power, 1890–1920’, in K. Deutsch, B. Fritsch, H. Jaguaribe and A. Markovits (eds) Problems of World Modeling: Political and Social Implications, pp. 169–205. Cambridge, MA: Ballinger.

Nord, D.P. (1986) ‘The Ironies of Communication Technology’, Clio (April): 12–16.

Organization for Economic Co-operation and Development, Working Party on Telecommunication and Information (1998) Internet Traffic Exchange: Developments and Policy. Paris: OECD.

Paisley, W. (1990) ‘An Oasis Where Many Paths Cross: The Improbable Cocitation Networks of a Multi-Discipline’, JASIS 41(6): 459–68.

Pirolli, P., J. Pitkow and R. Rao (1996) ‘Silk from a Sow’s Ear: Extracting Usable Structures from the Web’, CHI 96 Proceedings, 13–18 April.

Pitkow, J. (1998) ‘Summary of WWW Characterizations’, Computer Networks and ISDN Systems 30: 551–8.

Schiller, H. (1992) Mass Communications and American Empire, 2nd edn. Boulder, CO: Westview Press.

Schiller, H. (1995) ‘The Global Information Highway: Project for an Ungovernable World’, in J. Brook and I. Boal (eds) Resisting the Virtual Life: The Culture and Politics of Information, pp. 17–33. San Francisco, CA: City Lights.

Simmel, G. (1955) Conflict & The Web of Group-Affiliations. New York: The Free Press.

Smith, M. (1999) ‘Invisible Crowds in Cyberspace: Mapping the Social Structure of the Usenet’, in M. Smith and P. Kollock (eds) Communities in Cyberspace, pp. 195–219. London: Routledge.

Terveen, L. and W. Hill (1998) ‘Evaluating Emergent Collaboration on the Web’, Computer Supported Collective Work 98 Proceedings, pp. 355–62, Seattle, November.

Universal Postal Union (1997) Annual Report. Berne: UPU Communications Service. (Also available online, URL: http://www.upu.int)

Van Alstyne, M. and Brynjolfsson, E. (1997) ‘Electronic Communities: Global Village or Cyberbalkans?’, Sloan School of Management Working Papers, URL (consulted April 1999): http://web.mit.edu/marshall/www/papers/CyberBalkans.pdf

Wittaker, S., L. Terveen, W. Hill and L. Cherney (1998) ‘The Dynamics of Mass Interaction’, Computer Supported Collective Work 98 Proceedings, pp. 355–62, Seattle, November.

Woodruff, A., P. Aoki, E. Brewer, P. Gauthier and L. Rowe (1996) ‘An Investigation of Documents on the World Wide Web’, WWW5 Proceedings, Paris, France.

Yeo, G. (1995) ‘The Soul of Cyberspace’, New Perspectives Quarterly 12(4): 18–25.

Zook, M. (1998) 'The Web of Consumption: The Spatial Organization of the Internet Industry in the United States', paper presented at Tomorrow's Cities Today: Building for the Future, Pasadena, California, 5–8 November.

---

[1] I will refer here to both social and geographic limits as 'borders'. Drake (1993), among others, draws the distinction between 'borders', which are geographic and 'boundaries' which are administrative. The social borders I speak of here do not fall neatly into either of these categories, and I will use the single term to refer to both geographical and social structures.

[2] Were widespread geographic coding like that recommended by Farrell et al. (1994) available, the ability to make these kinds of analyses would be greatly improved.

[3] Though some efforts have been made in this direction (Bharat et al., 1998; Chakrabarti et al., 1998).

[4] Net Wizards (http://nw.com) server survey, January 1998. The gTLDs (.com, etc.) are included under the US, which leads to some overestimation (OECD, 1998: 45–6), but not by a significant amount (Arnum and Conti, 1998).

[5] For example, if the top level page of a site was at http://www.yahoo.com/potatoes/ harvesting.html, only pages with URLs beginning with http://www.yahoo.com/potatoes/ were included.

[6] Geocities alone claims a total of 1.4 million individual sites on hosts in the US and Japan.

[7] A site from the Tonga (.to) and a site from the Niue (.nu) were coded separately (both were US-based), as these registrations are used frequently by those outside the country.

[8] Initially four sets of these model texts were created, one each for English, French, Spanish, and German. However, this approach was found ineffective and both the page language and the subject category of pages containing languages other than English were coded separately with the help of native (and non-native) speakers.

[9] The impact of these links extend into economic and social spheres. Belgians, for example, recognize American brands on-line more often than Belgian brands, and are more likely to spend money on American sites as a result (see http://www.insites.be/persbericht.htm, consulted in April 1999).

[10] 'Other countries' includes only those with external links in sample: Norway, Mexico, Belgium, Ireland, Finland, South Africa, Brazil, Austria, New Zealand, Singapore, South Korea, Denmark, Slovania, Thailand, Venezuela, Poland, Turkey, Indonesia, Macau, Colombia, Dominica, Malaysia, Hong Kong, Kuwait, Micronesia, Bermuda, Czech Republic, Egypt, India, Greece, and Uruguay.

[11] Take, for example, a German site in the sample. A law firm that dealt primarily with matters having little to do with the international sphere included an English translation of their site 'after having recogniced [sic] that the Alta Vista translation client translates the german [sic] legal expression "Erbrecht" into "vomit" . . .' (http://www.afs-rechtsanwaelte.de/, 15 July 1998).

[12] Reno v. American Civil Liberties Union (1996), 521 US 844.