HUGY00457

Internet-Based Measurement

M. Dodge, University of Manchester, Manchester, UK
M. Zook, University of Kentucky, Lexington, KY, USA

Glossary

G0005 **Domain Name** A domain name is a unique identifier (e.g., nytimes.com or manchester.ac.uk) associated with an IP address that allows users to easily access a specific Internet resource such as a website.

G0010 **IP Address** Internet protocol (IP) addresses (e.g., 64.246.60.38) uniquely identify sites on the Internet and are necessary to ensure the delivery of traffic. They are little seen or used by typical users.

G0015 **Latency** The time (measured in milliseconds) that it takes to transmit and receive data between two nodes on the Internet.

G0020 **Ping** A network utility which sends test data to a target Internet site to determine whether it is 'live' and accepting data and reports the latency to the site.

G0025 **Screen Scraping** The use of a computer program to automatically collect data or data output of Internet-based resources – most often a web page.

G0030 **Topological Location** The location of sites on the Internet in terms of how they are connected to the rest of the network rather than a point defined by latitude and longitude.

G0035 **Traceroute** A network utility which maps out the path that data packets take between two hosts on the Internet, showing all of the intermediate nodes traversed, along with an indication of the speed of travel for each segment of the journey.

G0040 **Whois** A utility to search for the ownership details of a domain name.

P0005 Internet-based measurement is a set of methods that have been applied to quantitatively describe the structure, workload, and use of the Internet. They provide a practical means of doing a kind of virtual 'fieldwork' on the Internet using online tools and network monitoring techniques to gather fine-scale primary data, as opposed to relying on aggregate secondary data sources (such as government statistics). Typically, these methods use freely available software tools and web resources to explore the internal topology of Internet links and/or the external geography of network infrastructure, content production, and use. By measuring the operation of the Internet in terms of where things are produced and consumed, who owns them, and how data travels, researchers are able to critically engage and analyze the key network of the information age. In pedagogic terms, the openness of these techniques can help users of the Internet to transform themselves from passive consumers to more informed and active explorers of their world.

P0010 Internet-based measurement as a methodology for human geography is important and innovative in several respects. First, the Internet itself has several important geographical dimensions, and quantitative measurement techniques can provide unique data to analyze this. The focus has been primarily on mapping the material geography of the infrastructures of the Internet via automatic surveys of connected hardware and software services. Knowing where things are physically located is useful analytically because variations in spatial patterns provide researchers insight into underlying processes.

P0015 Second, 'ordinary' users can explore and measure the structure and operation of the Internet for themselves. This is because the Internet was purposefully designed as an open network that encourages active exploration and experimentation. This methodology does not require a large investment in expensive, specialized tools to obtain large and representative data samples. Many of the tools and techniques for Internet-based measurement already exist, having been created by engineers for the practical purposes of debugging network problems. These tools can be successfully leveraged to generate data useful for the context of social science research questions by providing tactical knowledge of the network that cannot be gained in any other way.

P0020 Using the Internet to measure and map itself is particularly useful for studies of the social geography of online interaction or the economic geography of website production where location is a key variable. For example, Internet-based measurement can reveal how territorial geographies of regulation and enforcement, particularly obscenity or libel laws, help shape the location of Internet activity. In addition, because the freedom to surf the web is not universal, Internet-based measurement methodologies are useful in identifying and evaluating state attempts to censor the dissemination or consumption of information by their inhabitants. Technically, censorship is often performed on behalf of a

national government by commercial internet service providers (ISPs) who block content reaching customers using a list of banned websites or domain names. Conceptually, this is the same process as software filters that parents can install on individual personal computers (PCs) to block access to inappropriate materials by their children. However, given the dynamic nature of the Internet, this type of censorship is often partial and requires authorities to continuously update the list of blocked sites. More insidious censorship can also be undertaken by search engine companies, who adapt their algorithms to give different, selective, results to users in particular countries. Crucially, users are not made aware that their results have been altered; for example, Google's agreement with the Chinese government means it alters key search results for Chinese users, including searches on Tiananmen Square, Tibetan independence, and Falon Gong.

P0025 Third, the Internet is a repository of a wide range of data which can be collected and cross-referenced to allow researchers to create databases that measure offline phenomenon such as fine-grained geographies of crime statistics or apartment listings. Geographic location (e.g., postal codes) is one of the most effective means of indexing data (Internet based or otherwise) since it enables linkages to a vast array of existing secondary data, such as demographic statistics from standard censuses. More fundamental, however, is the ability of empowered users or social movements to aggregate data spread across multiple web sources at relatively low cost to cast new light on long-standing problems.

S0005 Locating the Nodes of Internet

P0030 The ability to reliably determine the geographic location of the nodes of the Internet's infrastructure is the first step for Internet-based measurement. This task is challenging as the Internet was designed as a logical network that only 'knows' about topology (i.e., the location of connections capable of exchanging data) which can have little to do with physical location, defined by geographic coordinates. Thus, while the Internet has a robust and scalable system of unique locations (e.g., identifiers like Internet protocol (IP) addresses or domain names), these locations are not fixed to physical points on the ground or any particular position in the network. Moreover, because the Internet is a network of networks, rather than a homogeneous entity, the control of these location identifiers is decentralized and fluid. In short, 'no one owns the Internet' as a whole, and instead each component part is owned and operated by many different organizations and individuals. Consequently, no one institution has a synoptic view of the whole Internet and no one maintains a register of where all the components are physically located.

Determining the geographical location of components of the Internet is further complicated because different characteristics of a site's operation can be in different places. For example, there are five distinct kinds of geographical location which are important for fully characterizing a website (or other Internet resource). These types of locations include: P0035

1. lexical: a website is where its content refers, P0040
2. hardware: a website is where its hardware server is physically located, P0045
3. production: a website is where the author/maintainer responsible for it is located, P0050
4. ownership: a website is where the legal owner is located, and P0055
5. users: a website is where its users are located. P0060

In some cases, all five locations will be largely coincident geographically (e.g., a university's website). However, it is easy to imagine plausible scenarios in which a web page providing information on vacationing in Lexington, Kentucky, is hosted on a server in London, written by someone in Manchester for a website owner in Miami, which is read by people from across the world. The geographical precision of these different physical locations can also vary. Sometimes, location might be determined as the precise x, y position (e.g., street address of the building containing the web server); other times one might only know city or national jurisdiction. Each type of geographical location of a website is determined via different techniques. P0065

Geography of Content S0010

The first, and most obvious, method for determining the location of a website is based on lexical geography. Here, the content of the website is browsed to try to find an 'about page' or 'contacts page' that provides a postal address or telephone number for the website. Other cultural and linguistic clues (e.g., flags, symbols) in the content of a website might give useful indications of 'real-world' location. This method, however, is far from foolproof as many Internet resources do not provide readily identifiable measures of this type or do not allow a researcher access, for example, password-protected sites. Additionally, it is an extremely time-consuming approach as it requires a human being to visually inspect and categorize each site. P0070

Geography of Hardware S0015

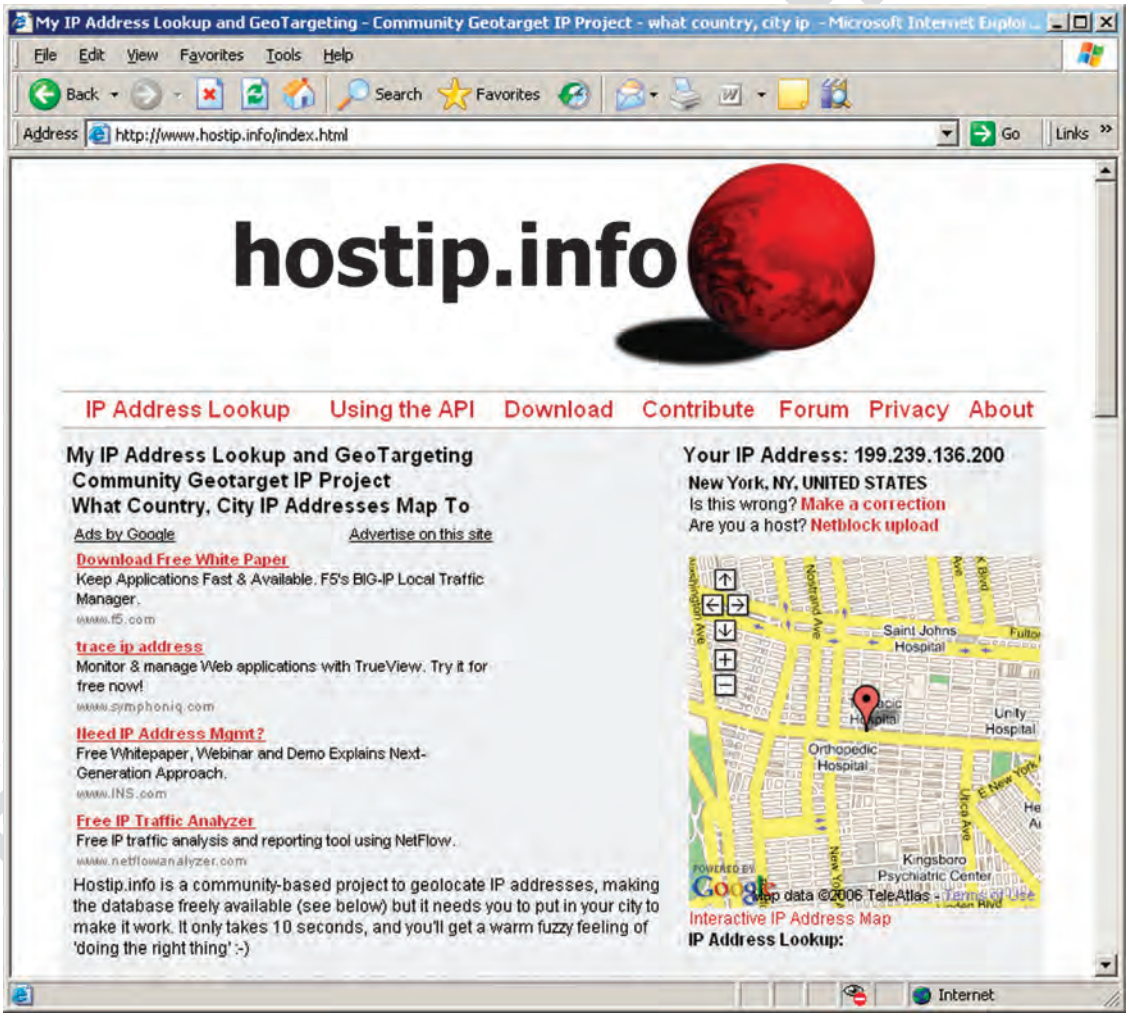
The second, and arguably most straightforward geographic measure of the Internet relies upon IP addresses. IP addresses are unique numeric identifiers, for example, 169.229.39.137, assigned to networked computers to P0075

exchange data. A variety of private and public databases exist that provide the associated geographic information for a particular IP address (**Figure 1**). While not foolproof, geo-coding IP addresses are reasonably accurate (particularly at the national level) and are widely used by companies to track users, guard against credit card fraud, and provide web content tailored for different territories. Moreover, it is possible to automate this process via software scripts in order to locate tens or hundreds of thousands of IP addresses in a very short amount of time.

P0080 The weakness of IP address geo-coding is that the use of anonymizers and other techniques can mask a user's actual location. More important for researchers is that most websites are hosted at dedicated server farms that have little to nothing to do with the location where the content for the site is generated or where the owner is located. Thus, IP addresses often highlight Internet infrastructure locations rather than content production centers. The applicability of this, of course, depends upon the research question pursued.

Geography of Production/Ownership S0020

P0085 Because IP addresses are awkward for people to use, the domain name system (e.g., nytimes.com or manchester.ac.uk) was introduced in the 1980s and now comprises a key component for Internet navigation and measurement. Domain names are organized according to top-level domains (TLDs) consisting of country code TLDs (ccTLDs) associated with domains ending with two-letter International Organization for Standardization (ISO) country code (e.g., .ca for Canadian domains, .ie for Irish domains, etc.) and generic TLDs (gTLDs) such as .com, .net, or .org. Approximately 35% of all domains are under ccTLDs and provide a crude measure of geographic location. However, the use of a country code domain name does not guarantee that the website is actually within the country indicated. The ownership, production, hosting, and use of that website



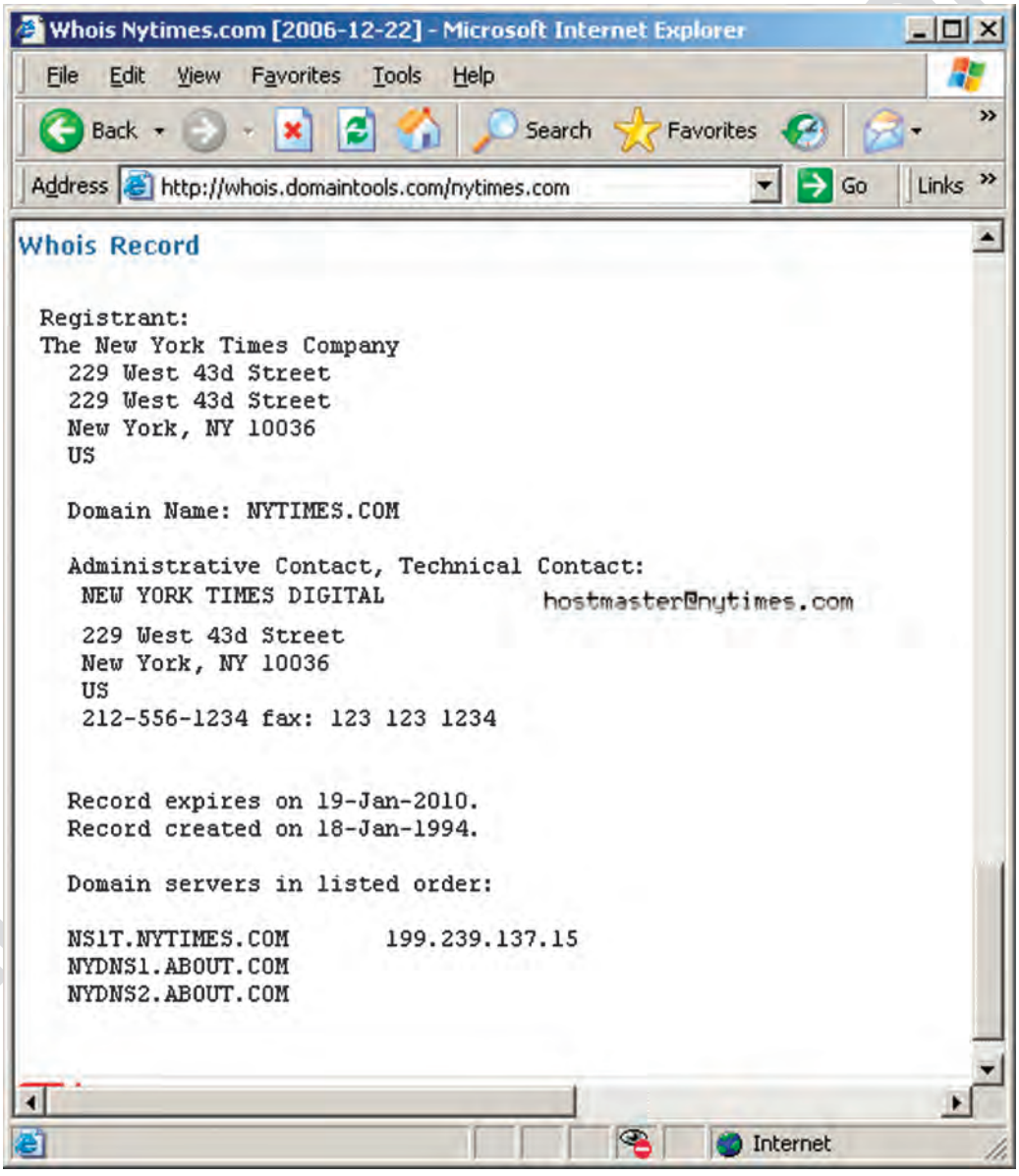
F0005 **Figure 1** The result of database lookup on an IP address using the web service offered by hostip.info that gives a geographic address for the registered owner. Source: author screenshot.

AU1

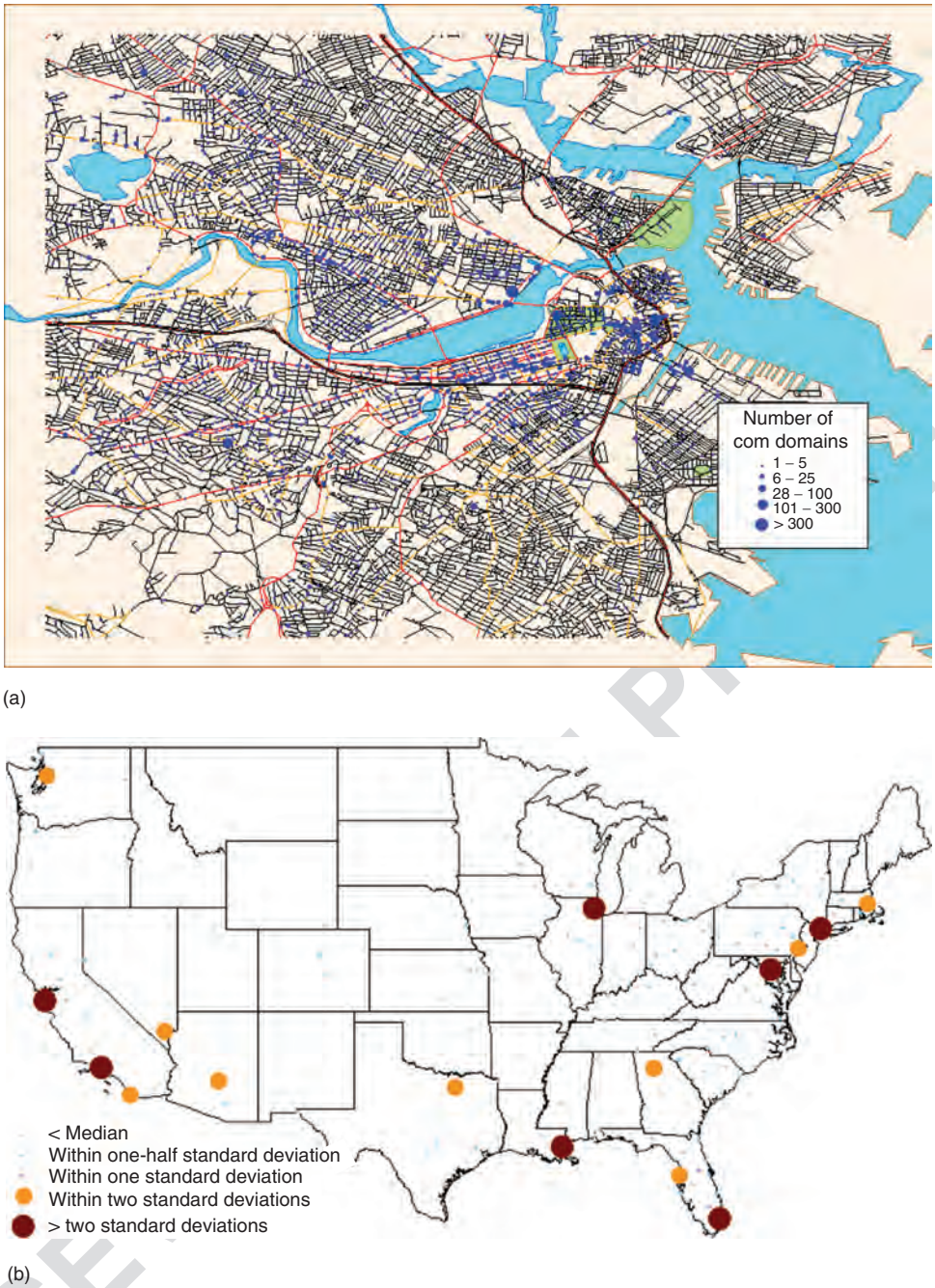
could well be in another country or several different countries. Furthermore, approximately 65% of domain names fall under the category of gTLDs and are not related to any country.

P0090 Thus, a much more accurate geographic location for domain names is derived via the online utility known as 'whois' which provides the ownership (listed as the registrant) information for a particular domain. Generally, it is possible to freely consult this registration information via a whois query but not all domain registration databases publicly give out the full address details of the owner. A whois query can be done interactively from any number of websites (**Figure 2**) and multiple whois queries can also be automated using software scripts.

While the results of whois queries can be helpful in finding out where the registered owner of a domain name is, they are not always accurate. First, registration details held on a given domain name may be out-of-date, incorrect, or deliberately false (e.g., spammers try to hide their true geographic location and would be unlikely to complete the registration honestly). Second, registrant information from a whois query only provides one location for a domain and it is not possible to determine whether this is indicative of the site of ownership, production, or both. Third, the registrations for large organizations often give a single postal address (their headquarters) and, thereby, may mask where the content P0095



F0010 **Figure 2** The result of a whois lookup on the nytimes.com domain name using a free web service called domaintools.com. The output gives registration details including the postal address of the owner. Source: author screenshot.



F0015 **Figure 3** (a) The ownership pattern of .com domain names in the Boston metropolitan region (July 1998) and (b) the US distribution of adult websites at the metropolitan statistical area (MSA) level (July 2001). The darker and larger circles indicate the MSA are more than two standard deviations above the average number of websites per MSA. Source: Matthew Zook.

for these individual domain names is actually being produced.

P0100 Despite these issues, the technique of using whois information for geo-coding websites were vital to the research by Matthew Zook on the geography of Internet content production. Using automated whois queries, he gathered comprehensive data on the location of .com domain name registrations in 1998 and demonstrated that the production of Internet content exhibited a significant degree of clustering in particular cities in the US and globally. Relying upon these techniques (supplemented with the use of IP location data), he has analyzed the geographies of a number of Internet-based activities ranging from the clustering of Internet startup companies during the 1990s to the location of adult-oriented websites (**Figure 3**).

S0025

Geography of Users

P0105 The location of users of websites is arguably the most difficult to measure as is the least centrally organized aspect of the Internet. Moreover, it is an ill-defined and dynamic variable as new people continuously come online and existing users adapt their online practices. At the level of individual websites, however, it is possible to gather rich data on the number, location, and activities of users (**Figure 4**). While potentially helpful, it can only shed light on the users of a particular website and gaining access to the user logs of leading websites is a difficult undertaking at best. For example, much could be learned by analyzing the geography of users of sites such as Google, Amazon, or eBay but these data are closely guarded as commercially sensitive.

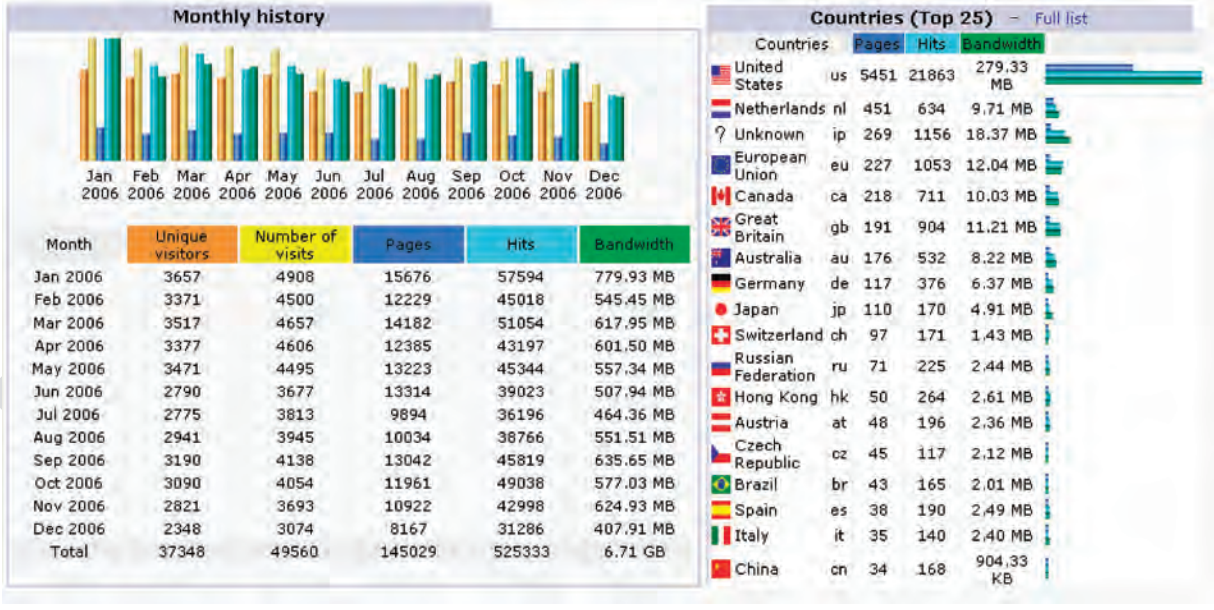
P0110 Measures of the number of Internet users at the national level is available but suffers from several drawbacks. First, the scale at which this data is organized prevents analysis at any subnational units such as city or region. Also problematic is that the data on users is generally constructed by combining national sample surveys employing different methodologies and definitions of Internet use. Finally, these data provide at best a measure of the potential demand for Internet resources but say nothing about the types of activity in which users are engaged. Although less comprehensive, researchers have generally conducted their own surveys of users to gain higher granularity and more specificity.

P0115 A middle ground between the rich albeit narrowly focused data from individual websites and the shallow yet comprehensive data from counts of users, are rankings that provide measures of the amount of user traffic to all websites. Although there are a number of ranking services, Alexa.com provides a long-running and independent measure of the popularity of websites among users. Alexa.com's rankings are based on tracking the surfing activity of a panel of Internet users who have downloaded a web browser tool. Alexa.com relies upon this sample to judge the most popular websites on the Internet and even disaggregates the top websites per country (**Figure 5**). While this represents a reasonable approach, it is unclear whether these users are a representative sample of Internet users (particularly when disaggregated to the country level) and several avenues for bias have been identified by Alexa.com and others. Nonetheless, it provides one of the best publicly available means to compare the location of the users of websites at the country level.

S0030

Measuring Distance and Routes Across the Internet

P0120 Another important element of Internet-based measurement is assessing distance between sites within the network. Given the topological structure of the Internet, physical distance between sites has little meaning. Instead, relative distances are measured using the journey time (i.e., latency) taken to transmit and receive data. Increasing latency implies increasing relative distance between two sites on the Internet. It is important to note, however, that there can be many different technical factors (e.g., types of hardware and network



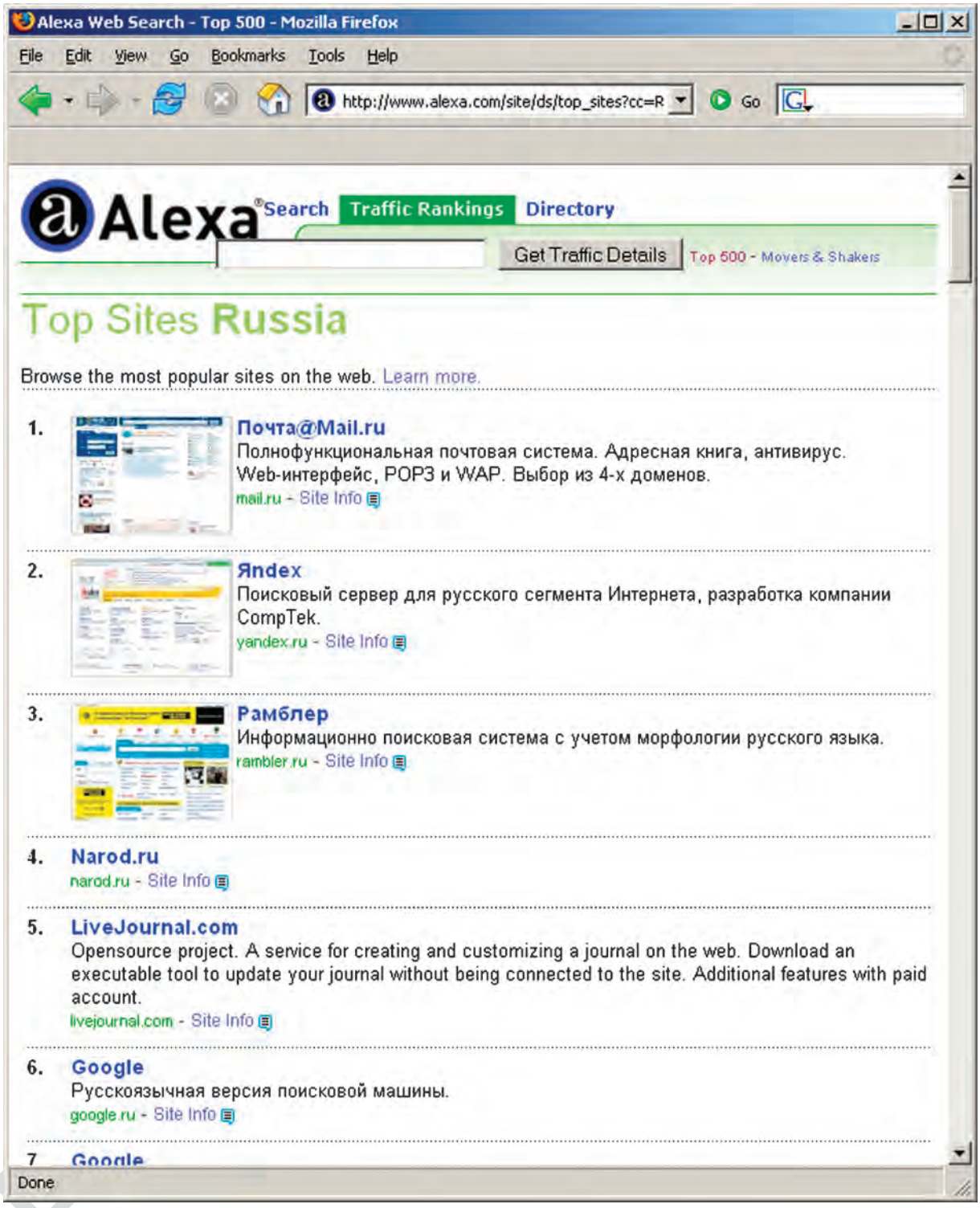
Monthly history

Month	Unique visitors	Number of visits	Pages	Hits	Bandwidth
Jan 2006	3657	4908	15676	57594	779.93 MB
Feb 2006	3371	4500	12229	45018	545.45 MB
Mar 2006	3517	4657	14182	51054	617.95 MB
Apr 2006	3377	4606	12385	43197	601.50 MB
May 2006	3471	4495	13223	45344	557.34 MB
Jun 2006	2790	3677	13314	39023	507.94 MB
Jul 2006	2775	3813	9894	36196	464.36 MB
Aug 2006	2941	3945	10034	38766	551.51 MB
Sep 2006	3190	4138	13042	45819	635.65 MB
Oct 2006	3090	4054	11961	49038	577.03 MB
Nov 2006	2821	3693	10922	42998	624.93 MB
Dec 2006	2348	3074	8167	31286	407.91 MB
Total	37348	49560	145029	525333	6.71 GB

Countries (Top 25) – Full list

Countries		Pages	Hits	Bandwidth
United States	us	5451	21863	279.33 MB
Netherlands	nl	451	634	9.71 MB
Unknown	ip	269	1156	18.37 MB
European Union	eu	227	1053	12.04 MB
Canada	ca	218	711	10.03 MB
Great Britain	gb	191	904	11.21 MB
Australia	au	176	532	8.22 MB
Germany	de	117	376	6.37 MB
Japan	jp	110	170	4.91 MB
Switzerland	ch	97	171	1.43 MB
Russian Federation	ru	71	225	2.44 MB
Hong Kong	hk	50	264	2.61 MB
Austria	at	48	196	2.36 MB
Czech Republic	cz	45	117	2.12 MB
Brazil	br	43	165	2.01 MB
Spain	es	38	190	2.49 MB
Italy	it	35	140	2.40 MB
China	cn	34	168	904.33 KB

F0020 **Figure 4** An example of typical website usage statistics giving in detail the number and nationality of visitors. Source: author screenshot.



F0025 **Figure 5** A sample of Alexa.com's data on website popularity including a listing of the most popular websites in Russia. Source: author screenshot.

configurations which are shaped by ownership and institutional structures) that effect latency. An interesting point of analysis, both on the Internet and in the 'real' world, is to compute the relationship between distance on the ground and time–distance for different places. This relationship is not always linear because of barriers, lack of connectivity, and poor accessibility. Sometimes, the quickest sites to reach are not the closest physically

but are 'institutionally close', while locations just down the street are completely offline. Analyzing the variable patterns in time accessibility can provide insight into underlying structural processes.

S0035 ## Distance Measurement

P0125 The simplest technique to measure latency uses the network utility 'ping' which reports whether a particular site on the Internet is 'live' and accepting data. It works by sending out test data to a target site and listening for a response. It is useful for distance measurement because it reports the round-trip time of data packets. For example, **Figure 6** shows the time (in milliseconds) each packet took to go from Lexington, KY to the web server at the University of Manchester and back again. The last line of the output reports the overall statistics. According to this, the average 'distance' for this particular journey across the Internet as measured by latency was 143 ms. Latency distances are very susceptible to changes in conditions on the Internet and can provide ways of quantifying possible traffic congestion, much like measuring car speeds gives an indication of the level of road congestion.

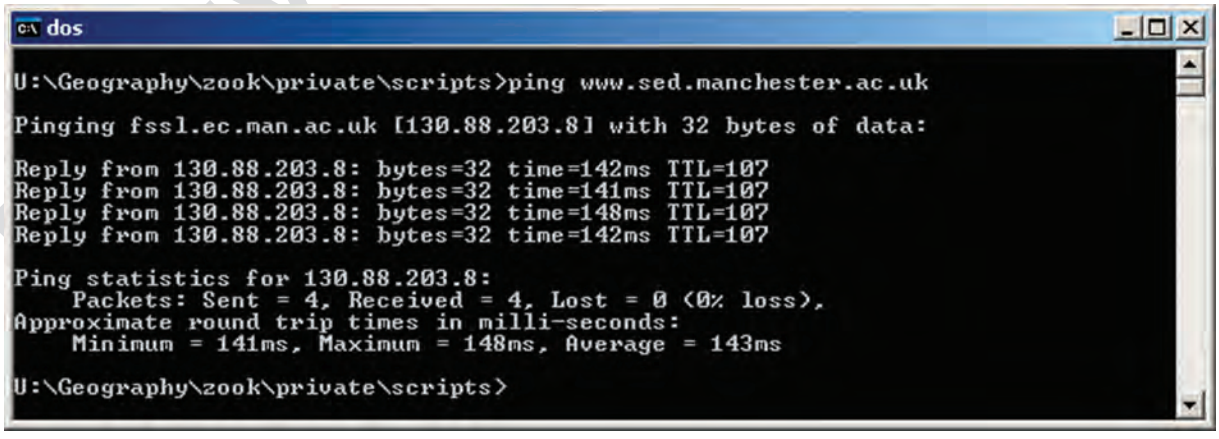
P0130 There also are several ways that 'pinging' latency distances can be used to learn more about structure of the Internet. First, and most obviously, a sequence of pings to the same site at different time periods can be used to build up a comprehensive longitudinal profile of latency. Another useful extension is to take pings from different places on the Internet to triangulate in on a particular site. By triangulating from different points it is possible to get a sense of the relationship between latency and physical distance, assuming that the (approximate) geographic location of the origins and target are known. More importantly, combining the latency and physical distance can provide a measure of whether a place is readily accessible on the Internet or not.

S0040 ## Data Route Measurement

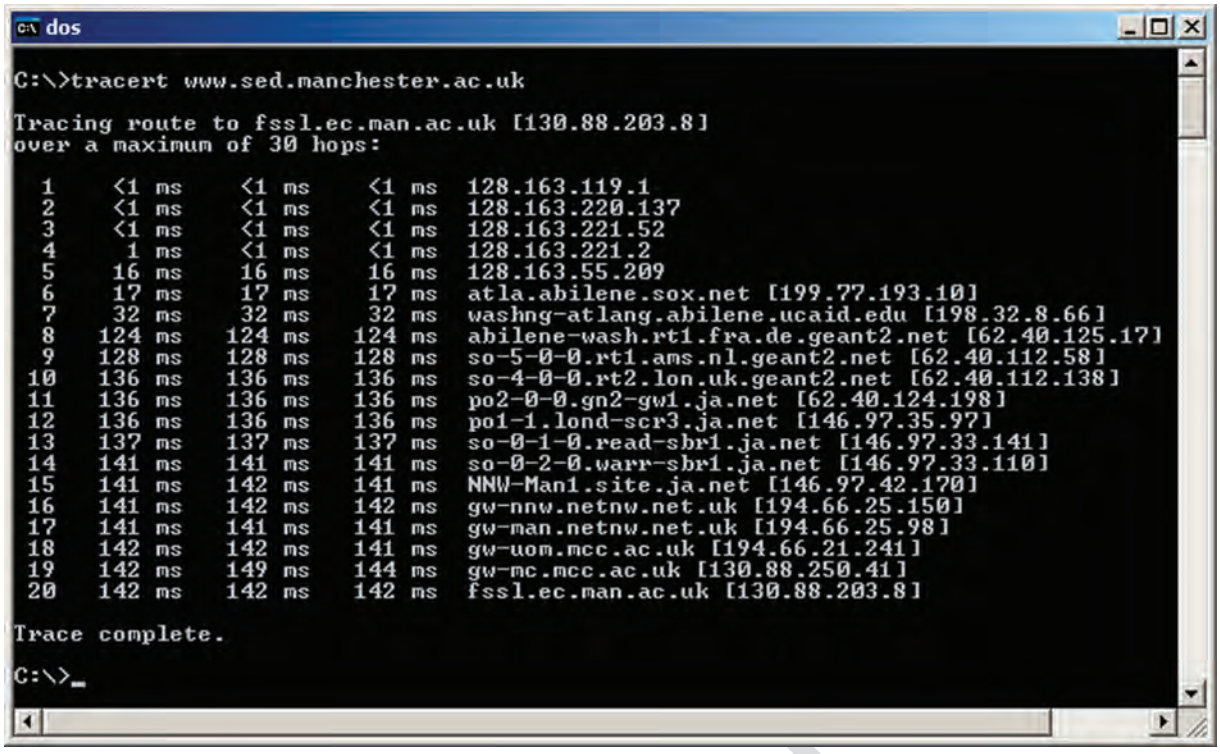
P0135 A much more sophisticated means of measuring distance through the Internet than ping is gained via the use of the utility 'traceroute' which reports details on the route data take through the Internet. Traceroute is invoked in much the same way as ping but provides greater detail. It effectively maps out the path that data packets take between two sites on the Internet, showing all of the intermediate nodes traversed, along with an indication of the speed of travel for each segment of the journey. Although traceroute is primarily for network engineers debugging routing problems, it has also been used by researchers to expose the political-economic structures of the Internet. It reveals the hidden complexity of data flows, showing how many nodes are involved, the seamless crossing of oceans and national borders, and the sometimes convoluted transfers through separate networks owned and operated by competing companies.

P0140 To illustrate how traceroute maps the Internet, it was used to chart the path from a PC at the University of Kentucky in Lexington, KY to a web server at the University of Manchester (**Figure 7**). The output looks rather cryptic at first sight, but it is in fact a kind of one-dimensional map, with each node traversed listed on a separate line. It gives a complete linear route listing showing how data packets traveled through the Internet starting in Lexington, traveling via Atlanta, Washington, Amsterdam, London, Reading, and Warrington and ending at Manchester. The three-time measurements in milliseconds – such as 142 ms, 149 ms, and 144 ms – are round-trip times for that segment and give a useful indication of the speed of each link.

P0145 Each node traversed is identified by its domain name and/or IP address. Most nodes have unusually long domain names (e.g., atla.abilene.sox.net) which are specialized routing computers at the core of the Internet not normally seen by users. With a little bit of decoding, the



F0030 **Figure 6** A ping query to a web server at the University of Manchester that provides a measure of latency between origin and target. Source: author screenshot.



F0035 **Figure 7** Traceroute measurement from the University of Kentucky to the University of Manchester. Source: author screenshot.

names of these routers can yield useful information, such as the type of node hardware, the bandwidth of the link, the name of the ISP that owns a node, and often a node's approximate location (usually at the city level). Many large network operators apply consistent naming conventions throughout their infrastructures, as in the machine names of the nodes of geant2.net. For example, the 'lon.uk' portion of the router name for segment 10 could AU3 reasonably be taken to mean London, UK (**Figure 8**).

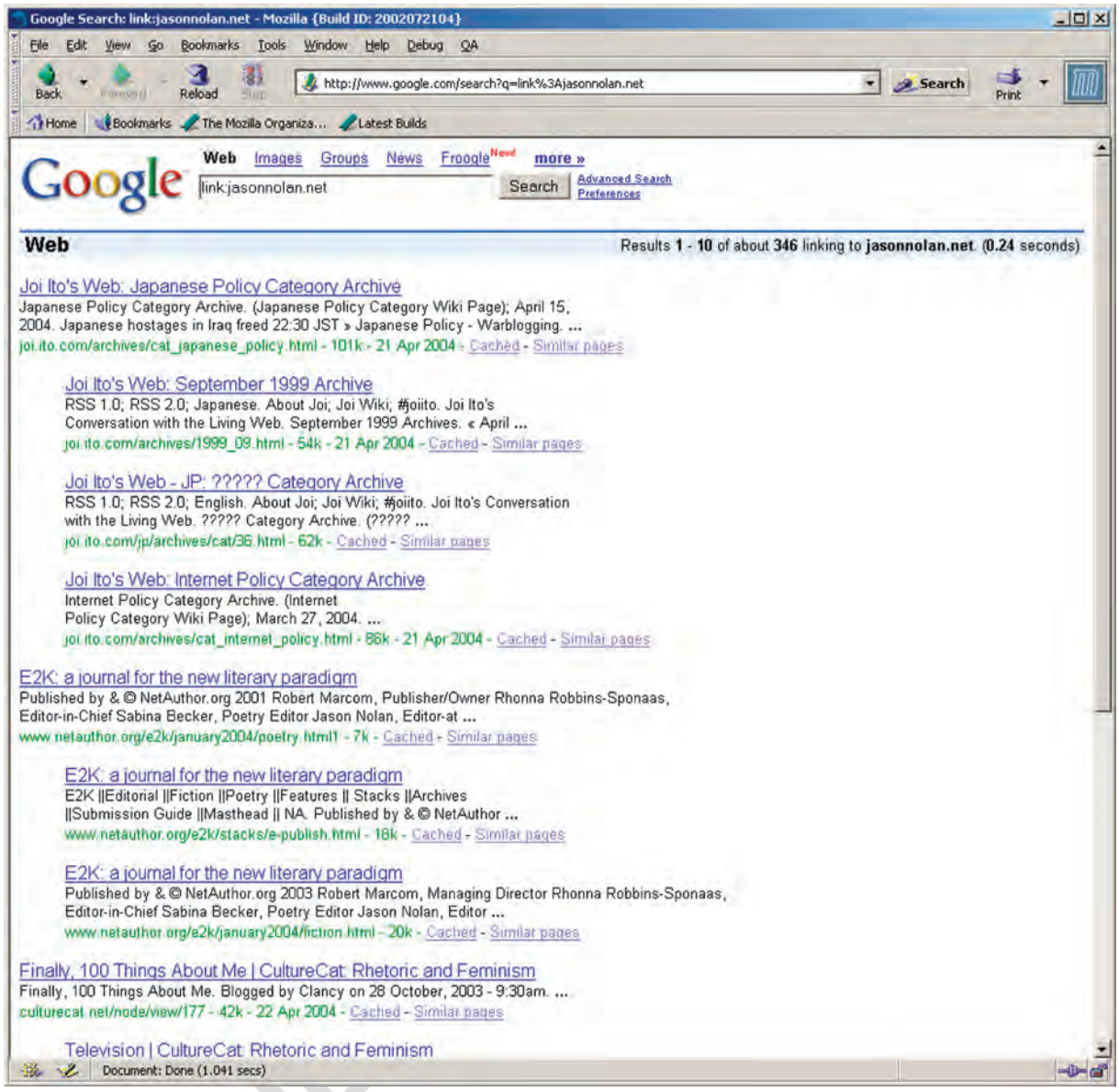
S0045 ## Utility of Traceroute Measurement in Research

P0150 Just like ping, the usefulness of single traceroutes can be extended by running them from different sites to triangulate the Internet's structure. Web-based traceroutes make it possible to run traces from many different starting points, including from different networks and in different countries. Web traceroute gateways are very useful for the active exploration of the Internet's topology from across the globe and illustrate the degree to which routes vary. Running multiple traceroutes to lots of different points across the Internet has also been used by researchers to gather large datasets on the topology of the core of Internet.

P0155 Data gathered by traceroute can provide evidence of the Internet's business 'logic' of following the cheapest paths rather than the shortest. Much international Internet traffic is still routed through the US as the cheapest means of transit between regions. This can result in quite anomalous looking, geographically circuitous routes being chosen.

P0160 Traceroute data can also be used for forensic analysis of the Internet's structure. For example, it is useful for deducing the approximate location of Internet hosts (such as websites) in terms of 'hardware geography'. The output shows the location and identity/owner of the 'upstream' network provider even if the final destination of the server is unclear. If the data travels into a certain city and does not leave it again, it is probable that the target is located there. Also, the 'upstream' network providers may keep logs for identifying a host that is of interest (this is of particular concern for law enforcement agencies in tracing the source of illegal activities).

P0165 Traceroute also reveals something of the hidden political economy of the Internet. The patterns of traffic routing show transit agreement and mutual peering relationship between competing companies. Details on these arrangements are often deemed commercially confidential, but are revealed by necessity in how and where the actual networks interconnect to share data. The routing of traffic reveals the structuring of business relationships in terms of who connects to who and the hierarchy of these connections (from periphery to center to periphery again). It can also show which telecommunications carriers dominate the transfer of traffic between particular countries and continents. These companies are likely to be influential in the structuring of



F0040 **Figure 8** .

AU2

global communications and tracerouting could provide an alternative way to quantify the extent of their power.

P0170 Last, the output from traceroute provides a useful way to assess the number of international borders crossed and determine which territories (i.e., separate legal jurisdictions) the data transits. The more 'points of contact' in the flow from origin to target, the more potential there is that Internet traffic could be intercepted and subjected to local regimes of monitoring, filtering, censorship, and data retention. In short, does an e-mail message transit through a third-party nation that has hostile intentions, for example, does an e-mail to someone in Palestine transit through Israel? Particularly in regions of conflict, being able to identifying territories that are transited might be vitally important in terms of the reliability of communication.

Internet-Based Measurement of Other Socioeconomic Phenomena

S0050

P0175 A final aspect of Internet-based measurement is its use as a means to gather data on phenomenon in physical space. One common method is the use of web-based surveys. While these surveys are relatively easy to setup and conduct, sampling design and response rates are crucial to ensure statistically meaningful results. Researchers must ensure that their sample accurately represents the population they are studying rather than simply being a

convenient way to conduct a survey. A related use of the Internet is to gather data that, particularly for datasets, are either prohibitively expensive to access or simply do not exist. A number of geographers have already leveraged the electronic format of the Internet to construct datasets (ranging from firms receiving venture capital to the location of corporate board members) in a relatively straightforward and, more importantly, cost-effective manner. To date, most of these efforts are largely manual retrieval operations driven by the fact that the desired data exists in multiple locations and does not have a standardized format.

P0180 An extension of this method known as screen-scraping allows access to data that would not otherwise be available or would be excessively time consuming to aggregate into a dataset. Screen-scraping is defined in this context as the use of a computer script (generally written by the researcher) to automatically collect the data output of Internet-based resources – most often web pages. These scripts are akin to user-written macros to automate simple but repetitive tasks in a spreadsheet and are designed to generate automatic queries to web pages and collect and store the data received. While it is possible to conduct the queries by hand, screen-scraping automates the process, greatly reducing research drudgery. It does, however, require programming and data-cleaning skills as there is no 'off-the-shelf' program that is readily available. Instead, researchers must craft their scripts specifically to the structure of the data which they are trying to collect.

P0185 A ready application of screen-scraping is the well-established technique of using directories as a method for identifying the locations of businesses. This method has been previously utilized with paper-based data within historical urban geography. Although issues of the reliability and accuracy of directories are a concern, the key limiting factor in the use of this method is the amount of work needed to assemble the data. The technique of screen-scraping can significantly simplify this task, making the collection of up-to-date, fine-scale spatially referenced data, (e.g., superfund sites, retail site locations) an easy task.

S0055 Conclusion

P0190 As the Internet grows in size, expands in scope, and becomes increasingly embedded as a banal background to everyday living, it becomes evermore important to understand the politics surrounding its production. Understanding the topological structures and geographies of the Internet, through quantitative network measurement using the techniques and tools described here, provides one of the most valuable avenues into network politics (e.g., the issue of net neutrality), allowing researchers to gather information firsthand and critically question network operations directly. The medium of communication might be virtual, but the Internet is dependent on physical infrastructure and human labor, most of which is invisible to users. The computers are small in scale and are usually hidden from view in anonymous server rooms and secure, windowless buildings, while the cables are under floors, in ceilings, and in conduits buried under roads. The technical geography of Internet infrastructures are easily overlooked (just like for other essential utilities of water, electricity), but they are not naturally given. The geographical structure and operation of networks that service modern living can be exposed through Internet-based measurement.

P0195 Internet-based measurement is likely to become easier as new and more powerful software tools for scanning the structure of Internet become available. Also, as commercially-provided search engine tools develop, they are increasingly providing new ways of surveying the information structures of the web. Of course, researchers will continue to have to tread carefully the ethical boundaries between critical fieldwork and potentially criminal hacking. At the same time, Internet-based measurement is also getting harder and riskier to do. Many parts of the Internet are being designed and operated in a much more closed fashion. For example, some networks block ping and traceroutes as a security precaution against malicious scanning. Although this makes Internet-based measurement more difficult, it becomes evermore important to ensure that researchers remain capable of analyzing it independently. In particular, as the Internet intertwines with physical space it is essential that human geography follows this evolution in order to understand the increasingly hybridized spaces inhabited in the twenty-first century.

See also: Cyberspace/cyberculture (00937); Georeferencing, geocoding (00448); Internet (00188); Mapping Cyberspace (00047).

Further Reading

Barabasi, A.-L. (2003). *Linked: The New Science of Networks*. New York: Perseus Books.

Branigan, S., Burch, H., Cheswick, B. and Wojcik, F. (2001). What can you do with traceroute? *Internet Computing* 5(5), 96.

Castells, M. (2001). *The Internet Galaxy*. Oxford: Oxford University Press.

Cukier, K. N. (1999). *Bandwidth Colonialism? The Implications of Internet Infrastructure on International e-commerce*. Paper presented at INET'99 Conference San Jose, California. http://www.isoc.org/inet99/proceedings/1e/1e_2.htm (accessed Mar. 2008).

Dodge, M. and Kitchin, R. (2000). *Mapping Cyberspace*. London: Routledge.

Dodge, M. and Kitchin, R. (2006). Net: Geography fieldwork frequently asked questions. In Weiss, J., Nolan, J., Trifonas, P., Nincic, V. &

Hunsinger, J. (eds.) *The International Handbook of Virtual Learning Environments*, pp 1143–1172. Netherlands: Springer.

Grubesic, T. H. (2002). Spatial dimensions of internet activity. *Telecommunications Policy* 26(7–8), 363–387.

Hayes, B. (1997). The infrastructure of the information infrastructure. *American Scientist* 85(30), 214–218.

Lakhina, A., Byers, J. W., Crovella, M. and Matta, I. (2002). *On the Geographic Location of Internet Resources*. Technical report 2002–15. Computer Science Department, Boston University. http://www.cs.bu.edu/techreports/pdf/2002-015-internet-geography.pdf (accessed Mar. 2008).

Murnion, S. and Healey, R. G. (1998). Modeling distance decay effects in web server information flows. *Geographical Analysis* 30(4), 285–303.

Shiode, N. and Dodge, M. (1999). Visualising the spatial pattern of internet address space in the United Kingdom. In Gittings, B. (ed.) *Innovations in GIS 6: Integrating Information Infrastructure with GI Technology*, pp 105–118. London: Taylor & Francis.

Spring, N., Wetherall, D. and Anderson, T. (2004). Reverse engineering the internet. *ACM SIGCOMM Computer Communication Review* 34(1), 3–8.

Townsend, A. (2001). Network cities and the global structure of the internet. *American Behavioral Scientist* 44(10), 1697–1715.

Zook, M. A. (2000). The web of production: The economic geography of commercial internet content production in the United States. *Environment and Planning A* 32, 411–426.

Zook, M. A. (2005). *The Geography of the Internet Industry*. Oxford: Blackwell.

Relevant Websites

http://www.alexa.com
Alexa the Web Information Company: website popularity survey.

http://irrepressible.info
Amenesty International.

http://www.arin.net
American Registry for Internet Numbers (ARIN): whois query interface.

http://www.domaintools.com
Domain Tools.

http://www.cia.gov
Central Intelligence Agency (CIA): World Factbook, Internet Users.

http://www.caida.org
Cooperative Association for Internet Data Analysis.

http://www.hostip.info
HostIP IP Geo-Coding.

http://www.iana.org
Internet Assigned Numbers Authority: IANA's list of all the country code top-level domains.

http://www.isc.org
Internet hosts survey, conducted by Internet Systems Consortium/Network Wizards.

http://www.lumeta.com
Internet mapping project.

http://www.internetworldstats.com
Internet Usage World Stats.

http://www.zooknic.com
Zooknic Internet Geography Project: Maps of Internet Users, Zooknic Internet Intelligence.

http://www.netcraft.com
Netcraft web server survey.

http://opennet.net
OpenNet Initiative.

http://www.rsf.org
Reports without Borders.

http://www.telegeography.com
TeleGeography.

http://www.traceroute.org
Thomas Kernen's web traceroute list.

http://www.visualroute.com
Visual Route.