Social Network Services as Data Sources and Platforms for e-Researching Social Networks*

Robert Ackland

Social network services such as Facebook provide new data for social science research into, for example, the role of individual characteristics in friendship formation and the diffusion of tastes in social networks. This paper assesses the potential of social network services for social science research in two ways. First, it is argued that social scientists conduct hyperlink analysis differently to applied physicists and researchers from the library and information sciences, and face constraints (relating to theory, methods and availability of appropriate tools) that are not encountered in the other disciplinary approaches. However, the constraints regarding theory and methods are less likely to be faced by researchers of online social networks, and for this reason, the rise of Facebook and other similar services is a potential boon for empirical social scientists interested in networks. The second part of the paper focuses specifically on the availability of research tools, and it is argued that social network services may eventually serve as e-Research platforms for delivering social network analysis tools.

Keywords: e-Research, hyperlink analysis, social network analysis, social science methods, online networks, social network services, Facebook

Word count: 6361

* Forthcoming in the *Social Science Computer Review* Special Issue on e-Social Science, Vol. 27, No. 4, Winter 2009.

Introduction

The Web, and the Internet on which it is built, support many types of networks that are of interest to empirical social scientists. Examples of such networks are: Newsgroups hosted on Usenet, where individuals form ties by responding to emails posted by other individuals; virtual worlds where the network ties between individuals reflect participation in joint activities; hyperlink networks between organizational websites or individual webpages therein, and blogsites; and networks of individuals in social network services (SNS) such as Facebook, where the ties are “friendships” (where a user requests and gains permission to list another user as a friend on his or her profile) or joint membership in groups (for example, for alumni of particular universities). The present paper focuses on two particular types of online networks: hyperlink networks and networks enabled by SNSs (which are referred to here as "online social networks").[1]

There are two parts to this paper. First, while there is currently a lot of "buzz" about the opportunity presented by SNSs such as Facebook for empirical social science, social scientists have in fact been studying networks on the Web since the early days of Web 1.0. It therefore seems an appropriate time to relate current research into online social networks to on-going social science research into hyperlink networks. This is done by providing a unique typology of disciplinary research into hyperlink networks. In particular, it is argued that the social science approach to researching hyperlink networks is inherently different to that conducted by applied physicists and researchers from the library and information sciences (LIS). It is further contended that differences across the disciplines on three dimensions (theory, methods and availability of appropriate tools) provides an explanation for why empirical hyperlink analysis has had a slower uptake within the social sciences. The question then asked is: in the context of these three dimensions, how does research into online social networks differ to social science hyperlink analysis? It is argued that constraints regarding theory and methods are less likely to be faced by researchers of online social

networks, and for this reason, the rise of Facebook and other similar services is a potential boon for empirical social scientists interested in networks.

The second part of the paper focuses specifically on the issue of tools for social science research into online social networks, and in particular, how SNSs may serve as e-Research platforms for delivering tools and data to social scientists. To illustrate the potential of SNSs for e-Social Science, a prototype social network analysis tool (VOSON-os) delivered on the OpenSocial platform is presented.

Empirical research into online networks

There are various types of Internet-enabled networks that are of interest to empirical social scientists. Some examples are:

- Newsgroups - repositories of emails set up for different topics, often hosted on the Usenet system. In such a network, the nodes are individuals who form ties by sending emails that respond to previous emails (thus forming a "thread"). Discussion groups hosted on the Web are functionally similar to newsgroups (which do not necessarily involve Web technologies) and can be researched in an analogous manner. See, for example, Welser et al. (2007).
- Virtual worlds - computer-based simulated environments where individuals can assume digital representations (avatars) and interact with other individuals. Second Life is a popular example of a virtual world, and World of Warcraft is another prominent example of a particular subset of virtual worlds, called massively multiplayer online role-playing games (MMORPGs). With virtual worlds, the network nodes are individuals and the ties can be inferred by, for example participation in joint activities such as "raids" or "quests", or physical proximity in the world (e.g., Contractor et al., 2008).

- Hyperlink networks constructed by using web crawlers (or APIs such as those offered by Google and Yahoo) to identify hyperlinks between web pages. Researchers have studied hyperlink networks where the nodes are websites of organizations (e.g. Shumate and Dewitt, 2008) and also where the nodes are blogsites (chronologically updated websites, generally authored by a single person, with a "journal" type look and feel), see for example, Park and Jankowski (2008).
- Online social networks enabled by social network services such as Facebook, where the ties are friendships or joint membership in groups (e.g., Wimmer and Lewis, 2008).

The focus of the present paper is hyperlink networks and online social networks; the opportunities and challenges for social science research into each type of network are now discussed.

Hyperlink networks

There are three broad approaches for empirical research into hyperlink networks, each reflecting a particular disciplinary base.[2] First, webmetrics (also known as webometrics and cybermetrics) is an example of informetrics, a subfield of library and information science (LIS) which involves the use of mathematical-statistical approaches for the analysis of communication in science. Examples of webmetrics are Almind and Ingwersen (1997), Björneborn and Ingwersen (2004) and Thelwall et al. (2005). In a recent webmetric study, Barjak and Thelwall (2008) regress counts of inbound hyperlinks to the websites of life science research teams on relevant offline characteristics of the teams (e.g. gender of team leader, industry connections, research productivity) in order to assess the role of hyperlinks as science and technology output indicators.

Second, applied physicists have focused on the identification of empirical properties in large-scale collections of Web pages and the development of statistical-mechanical models that can be used to explain these properties.[3] For example, Newman (2002) studied the existence of “assortative mixing” (or correlation between the attributes of adjacent network nodes) in several large-scale

networks and found that the Web in fact exhibits *disassortative* mixing (or degree anti-correlation between nodes), reflecting the fact that high-degree web pages are often directories that by definition tend to connect to low-degree individual web pages. Barabási and Albert (1999) explain "power laws" in the distribution of links in networks such as the Web, where a small number of sites receive the lion's share of links pointing toward them, via the concept of preferential attachment: newer entrants are inclined to link to already well-connected actors, thereby increasing the incumbents' advantage and leading to a winner-takes-all situation.

Finally, social science approaches to hyperlink analysis have attempted to understand the role of the Web in enabling various forms of social, economic or political behavior An obvious aspect that distinguishes social science hyperlink analysis is a focus on actors traditionally studied by social scientists, for example, political parties (Ackland and Gibson, 2004), environmental social movement organizations (Ackland, O'Neil, Bimber, Gibson, and Ward, 2006), and civil society actors (González-Bailó, 2007). Social science hyperlink analysis also involves testing of hypotheses emerging from social science models. For example, Shumate and Dewitt (2008) use exponential random graph modeling techniques to identify "structural signatures" of collective action behavior in hyperlink networks, whereby HIV/AIDS NGOs are creating an "information public good". Compared with researchers from LIS and applied physics, social scientists are more inclined to view the Web as a relational space and this leads to social network analysis (SNA) being used to analyze how the structural position of actors impacts on actions and opportunities.

A major contention of this paper is that empirical hyperlink analysis has been slower to develop and gain widespread acceptance within the social sciences, compared with LIS and applied physics. This statement is not based on a rigorous bibliometric analysis of relevant research output within the various disciplines, but rather, reflects the experience and observation of a social scientist who has been researching hyperlink networks since 2002. Today, it can be argued that there is only limited acceptance from the traditional social science disciplines (economics,

sociology and political science) to the idea that hyperlink data can provide useful and meaningful insights into economic, sociological and political phenomena.

With regards to economics, the lack of general interest in hyperlink networks is not unexpected since networks are not part of the mainstream (economists generally view interactions between individuals and other actors as occurring through markets, rather than networks).[4] Similarly, network analysis has only recently become widely viewed as a useful addition to the toolkit of empirical political scientists, with the interest perhaps being largely driven by the growing importance of the Internet to political processes, especially in the US.[5] However, the lack of engagement with hyperlink network research from mainstream sociologists is unexpected since network analysis is already well established in many parts of sociology. For example, network methods are well-established in sociological research into social movements (see, for example, Diani 2001, 2003) and the field of social network analysis (the canonical reference being Wasserman and Faust, 1994) has developed within sociology (but is increasingly used in other disciplines).

Other than the issues for particular social science disciplines just identified, we propose three main reasons for the slow uptake of hyperlink network research in the social sciences in general, and sociology in particular. The first relates to the conceptual framework for studying the Web, and the theoretical meaning of hyperlinks and network nodes. An applied physicist's view of the Web as a large-scale network of hyperlinked documents is non-controversial in that this is simply a statement about the underlying architecture of the Web. Applied physicists are generally not concerned with attributing theoretical or behavioral meaning to hyperlinks, and focus rather on developing sophisticated approaches for characterizing and simulating hyperlink networks.

Researchers from LIS are more likely (compared with applied physicists) to be interested in the behavioral foundations of hyperlinks. In contrast with social scientists, LIS researchers have more readily embraced hyperlinks as having theoretical meaning, possibly because they can be seen as being analogous to citation networks traditionally studied within this field. To the extent that LIS

researchers study scholarly activity via hyperlink analysis, the concept of network node is more complicated than the individual web pages studied by applied physicists. However, there are well-established LIS approaches for aggregating web pages up to appropriate units of analysis (e.g. research teams, universities) using "alternative document models" based upon directories, domains and multi-domain sites (see, for example, Thelwall 2002, 2004).

A social science view of the Web is of a network of people and organizations, rather than a collection of hyperlinked documents. This world view leads to greater difficulty in developing conceptual frameworks for hyperlink networks; in the context of research into online social movements, for example, what does a hyperlink mean, and what are the appropriate network nodes? The potential for using SNA to analyze hyperlink networks was first noted in the relatively early days of the Web. Jackson (1997) considered that SNA "...has significant potential to generate insight into the communicative nature of Web structures" but was not comfortable with nodes in a hyperlink network (pages or sites) being described as social actors and further, argued that two of the core assumptions of SNA, the dependence of nodes within a network and the emergent property of networks, do not apply to the Web.

The second reason why social scientists have been slower to embrace hyperlink analysis relates to research methods, and in particular, the challenge of dealing with datasets that potentially consist of millions of web pages. The scale of such data does not pose a problem for applied physicists since they rarely require information on network nodes and ties other than that which can be obtained via automatic means (in particular, there is a focus on network attributes that are derived from the graph structure itself, e.g. indegree or outdegree, rather than attributes that pertain to a web page, web site or the owner of the website). While LIS researchers are interested in obtaining attribute data for websites representing, for example, research teams, webmetrics typically does not involve the use of complete networks (a complete network is where there is information on the outbound links made by all actors in the network). Barjak and Thelwall's (2008) webmetric analysis involved the use of total counts of hyperlinks pointing to a given sample of websites and

there was no need to attach attributes to the *senders* of these hyperlinks (in webmetrics, the focus is rather on the attributes of the *receivers* of hyperlinks).

Social science hyperlink research (in particular, where SNA is used) involves complete networks, so one either needs a small and well-defined population of sites obtained from a sampling frame that is external to the network (e.g Ackland and Gibson, 2004, study political party hyperlink networks) or else an approach for constructing network samples is needed. Sampling is a cornerstone of empirical social science research but it is not currently used in hyperlink analysis. In fact, there is potentially some resistance from computer scientists to the idea that limitations of automated data collection may necessitate sampling of Web networks, since this goes against the concept of the Semantic Web. The Semantic Web involves building into web pages tags for data, and semantic representations of the meaning of those tags, thus promising a Web that is machine readable (Berners-Lee et al., 2001; Shadbolt et al., 2006).[6]

A final reason why hyperlink analysis has developed more slowly in the social sciences is a lack of appropriate tools. Tool development is better accepted as a part of scientific process in applied physics and LIS (and researchers often have the skills to develop their own tools). For social science research into the Web there needs to be access to a set of heterogeneous tools that can enable observation and tagging of hyperlinks and/or network actors, and aggregation of pages into meaningful groups. While these tools might be adopted from other approaches, they need to be modified to support a social science view of the Web. It is unreasonable to expect that a tool that has been developed for an entirely different research context will make a big impact on social science Web research without major modification. A "Google for Social Scientists", with focus on full-text search over Web pages as the core research function, and no regard for sampling and statistical inference, is unlikely to have a big impact in social science research into the Web.

Online social networks

Research into online social networks is still in its early stages, and it is therefore difficult to assess the extent to which social scientists will embrace this new area (compared with research into hyperlink networks). However, social scientists appear to be well placed to engage with and undertake analysis of online social networks. Sociologists have been studying friendship networks for decades and have found that patterns of friendship are strongly affected by characteristics such as age, race and language - this is the "birds of a feather flock together" phenomenon (see, for example, McPherson et al., 2001). Online social networks offer new data for research into homophily (see, for example, Wimmer and Lewis, 2008) and also research into the diffusion of tastes.

To what extent are the three constraints identified with respect to social science research into hyperlink networks (difficulty of developing conceptual frameworks, methodological issues, availability of appropriate tools) likely to be different with online social networks? The first two constraints (conceptual frameworks and methods) are discussed here, while the third constraint (tools) is discussed in the next section.

It is arguably easier to develop a conceptual framework for social science research into online social networks. With SNSs, the nodes are clearly people and the network ties are also relatively easily interpreted. While it is obvious that a friendship in Facebook means something different to an offline friendship (in terms of the cost of making and maintaining the tie, for example, and in terms of the public-private nature of the action), it is perhaps easier to interpret a tie in Facebook compared with a hyperlink tie.

The relatively easy interpretation of network nodes and ties also means that social science research into online social networks is less methodologically challenging, compared with hyperlink networks. Users of SNSs are encouraged to describe themselves (e.g. political persuasion, religion, humor, smoking and drinking behavior) using a combination of text fields, drop-down selection

boxes and check boxes. The profiles are highly amenable to automated data analysis (compared with web pages, which are much less structured). However, there is of course an issue with data quality; SNS users may in some circumstances be inclined to exaggerate certain aspects of their profiles, and these data might be difficult to validate in the absence of follow-up surveys.

However, despite the machine-readability of SNSs, there are two reasons to expect that network sampling will still be important for social science research into online social networks. First, SNA techniques such as exponential random graph modeling currently have difficulty with large-scale networks. Further, these techniques are not well-suited to networks where a person can have 5000 "friends", and there will be a need for further data interrogation (perhaps viewing of the profiles to establish friendship using a different metric or else by directly asking the user to identify their "real" friends). Also, it is possible that individual attribute data needed by social scientists will not be contained on publicly-viewable user profiles (either because it isn't of interest to the creator of the SNS or else it is of a personal nature), and hence will need to be obtained via follow-up surveys. Recent advances in adaptive sampling (see, for example, Thompson 2006), which is typically used to sample from small or hidden populations (snowball sampling is an example of adaptive sampling), may be highly relevant for social science research into online social networks.

Social network services and e-Research

e-Research is the use of advanced ICTs to enable new forms of collaborative research that involves access to distributed research resources (datasets, methods, compute cycles). The terms e-Research and e-Social Science (which is e-Research in the social sciences) are sometimes misused to refer to research into online networks. It is useful to clarify that research into online networks is *not* the same thing as e-Social Science: e-Social Science refers to a mode of collaborative research involving advanced ICTs that often has nothing to do with the Web as an object of research or source of data. For example, the MoSeS project which has been funded as part of the UK e-Social

Science programme and aims to "develop representation of the entire UK population as individuals and households, together with a package of modeling tools which allows specific research and policy questions to be addressed" (http://www.ncess.ac.uk/research/geographic/moses/) is an example of e-Social Science which clearly has nothing to do with research into online networks.

Similarly, it is possible to conduct empirical social science research into online networks and not be doing e-Social Science. For example, using a web crawler such as Mike Thelwall's SocSciBot[7] to collect data on organizational hyperlink networks on the Web, with the data subsequently being analyzed using an SNA tool such as Pajek[8] would be an example of online network research, but not (using the above definition) e-Social Science, since the technologies involved do not enable collaborative access to distributed research resources. Unlike SocSciBot, which is client software, IssueCrawler[9] is a hosted web service (i.e. access is provided via a web browser) and thus it clearly facilitates access to distributed research resources. However, to our best knowledge, IssueCrawler does not enable *collaborative* access to these resources (that is, researchers who are not co-located being able to collaboratively work with common datasets and methods), and IssueCrawler is therefore not an e-Research tool, using the above formal definition. In contrast, the Virtual Observatory for the Study of Online Networks (VOSON), which is a web-based research tool for social science hyperlink analysis, has been specifically designed by the author to enable collaborative access to distributed research resources, and is thus an example of an e-Research tool for researching online networks.[10]

It was argued above that a limiting factor for social science hyperlink research is the availability of appropriate tools. It is proposed here that SNSs can, in addition to providing new data for social science research, serve as delivery platforms for e-Research tools that are specifically catered to social scientists studying online social networks.

VOSON-os - a demonstrator OpenSocial e-Research tool

In order to assess the viability of SNSs as platforms for e-Research tools, a prototype of the VOSON e-Research tool was developed using the OpenSocial API, using the Orkut SNS as the delivery platform. OpenSocial has been developed as a direct challenge to Facebook (who pioneered applications for SNSs).[11] The technical details of developing an OpenSocial application are not covered here (detailed tutorials are available online), but the VOSON-os prototype (Figure 1) was very easy to build (especially when compared with the challenges of working with portal development frameworks such as Gridsphere[12]).

VOSON-os features an interactive network map which is provided using JSViz.[13] The intention is that this map would show the ego network of the person who has installed VOSON-os (an ego network consists of a focal node ("ego") and the nodes directly connected to ego ("alters") plus any ties among the alters).[14] The OpenSocial API enables the programmatic identification of the friends of the user who has loaded the VOSON-os, but in order to know the ties amongst the friends of the user, it is necessary that those users also install VOSON-os. So, we can imagine a research strategy that might involve emailing people in a particular target group and asking them to install VOSON-os, and this would enable the construction of a complete network containing all users who had loaded the application (there would of course be a response rate less than 100%, but this is the same with SNA conducted offline). The application could also be used to elicit further information from the participants. For example, users of an SNS could be periodically asked to provide information on labor market status, which could then be correlated with professional networking behavior

VOSON-os also features basic graph-level SNA metrics calculated for the dummy data.[15] A Facebook or OpenSocial e-Research SNA tool would preferably make use of SNA libraries such as the sna library for the R statistical software[16] or Jung[17]. Ideally, these routines could be accessed via web services (however, the cross-domain security model in OpenSocial currently prevents this).

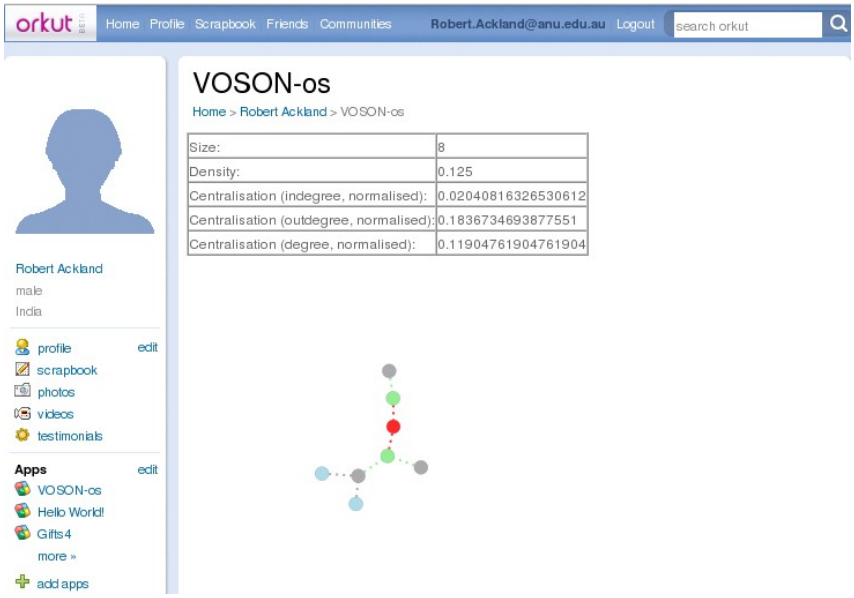

Figure 1. VOSON-os - a demonstrator SNA tool on the OpenSocial platform

The promise of e-Research for social science

Social scientists have been analyzing surveys of individuals and households for decades and are well served by existing tools such as Stata, SAS and SPSS. It is in new areas of research, such as research into online networks, that e-Research can make a major contribution to social science, by enabling access to new forms of data and research methods. Social science research into the Web requires diverse tools (e.g. web crawler, text mining, data visualization, SNA) that are unlikely to be provided by a single tool developer (indeed, such "vertical integration" could potentially lead to anti-competitive behavior).

e-Research promises the seamless connection (via web services) of research resources (data, methods, computational) from different providers. However, there have been problems in how this has been implemented. First, there has been a reliance on heavyweight middleware software such as Globus and portlet development frameworks such as Gridsphere (with the emphasis on JSR-168 compliance and user interfaces that are primarily designed to support batch submission of code, which is a mode of research which more suited to computational scientists rather than social scientists), eschewing simpler, but effective web technologies such as AJAX-enabled websites. The potential limitations of Grid middleware such as Globus for e-Science were pointed out by authors

such as Chin and Coveney (2004), and Ackland et al. (2006) emphasized the importance of AJAX-enabled web applications, featuring menu-based user interfaces (which are commonly used in social science research applications) for e-Social Science. Recently, the UK e-Research community have increasingly focused on the use of Web 2.0/Wiki technologies for building distributed collaboration tools, with MyExperiment[18] being the most notable example.

A second possible problem with regards to the implementation of e-Research in the social sciences arises from the roles of technologists and social scientists in the projects. The central involvement of technologists in e-Social Science (almost an inevitable consequence of the use of sophisticated technologies such as Globus and Gridsphere) has perhaps led to indirect pressure on social scientists to adopt research approaches from other disciplines, regardless of the fact that social scientists already have well-developed and sophisticated research frameworks. The result has been an emphasis on searching over digital research collections (something more important to the humanities than the social sciences), workflows and ontologies, while other important aspects of empirical social science (e.g. sampling and statistical inference) have been relatively neglected.[19]

There needs to be balance between the aims of technologists who must push technology boundaries to receive kudos and the needs of social scientists, who are often just looking for something “quick and dirty” that will enable them to decide whether to invest further in e-Social Science. A social scientist will gain kudos from publications in appropriate journals and the development of research tools is generally just a means to an end, rather than the end in itself.

It has been argued here that SNSs such as Facebook and Orkut present interesting possibilities as platforms for e-Research tools. The platforms are easy to develop on and while certain aspects (e.g. inter-application communication) are not currently possible, inter-portlet communication is not presently available in Gridsphere either (see, for example, Yang et al. 2006).

Further, it is contended that the OpenSocial API may prove to be an enabling technology that will facilitate the proliferation of (borrowing the terminology introduced by Benkler, 2006) "peer-

produced" tools and data for the advancement of social science research into online networks. The phenomenon of Internet-enabled collaboration which Tapscott and Williams (2007) term "wikinomics" and von Hippel (2005) refers to as "user innovation" has already made a large impact in particular sectors, with open source software development being the primary example. Social science research into online networks involves the use of a number of tools (e.g. web crawler, text mining, social network analysis) and it is not feasible for a single tool developer to be able to "keep up" with the modifications to the tools that are needed to keep pace with developments in research methodology. It is argued here that SNSs such as Facebook and Orkut can provide choice of tools and facilitate innovative user-driven e-Research into online social networks.

Conclusions

One of the motivations for writing this paper has been the frustration that the author has experienced when encountering (what he contends is) misuse of terminology pertaining to e-Research into online networks. This is clearly an emergent field, and the language to describe activities within the field is still being contested by various interested parties. This paper represents an attempt to clarify some of the issues and language surrounding the use of e-Research technologies for online network research.

This was done via a discussion of the impact of the phenomenal recent growth of social network services such as Facebook for social science research into online networks. The paper argued that the rise of Facebook and similar sites has important implications for social science online network research. First, online social network services enable a form of online networking that may assuage or minimize some of the theoretical and methodological constraints that have worked against widespread acceptance of Web data for social science research. Second, it was argued that these services might also significantly impact on the practice of empirical social science, providing new

e-Research platforms for user-driven innovation in tool development. MyExperiment, a distributed collaboration tool built using Web 2.0/Wiki technologies, has been described as "Facebook for scientists"; this paper contends that that Facebook and other social network services may eventually be described as "GridSphere for e-Social Scientists".

In conclusion, the emergence of Facebook and other social network services is a potential boon for social scientists, both in terms of new data for researching important social science phenomenon, and new tools for social science research.

Author details

Robert Ackland is a Fellow in the Australian Demographic and Social Research Institute, The Australian National University. Email: robert.ackland@anu.edu.au

References

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

Ackland, R., O'Neil M., Bimber B., Gibson, R. and S. Ward. (2006): "New Methods for Studying Online Environmental-Activist Networks," paper presented to 26th International Sunbelt Social Network Conference, 24-30 April, Vancouver.

Ackland, R., O'Neil M., Standish, R. and M. Buchhorn (2006): "VOSON: A Web Services Approach for Facilitating Research into Online Networks," refereed paper presented at the Second International Conference on e-Social Science, 28-30 June 2006, University of Manchester.

Almind, T. C., and P. Ingwersen (1997): "Informetric analyses on the World Wide Web: Methodological approaches to `webometrics'," *Journal of Documentation*, 55(4), 404-426.

Barabási, A.-L., and R. Albert. (1999): "Emergence of Scaling in Random Networks," *Science*, 286, 509-512.

Barjak, F. and M. Thelwall. (2008): "A statistical analysis of the web presences of European life sciences research teams," *Journal of the American Society for Information Science and Technology,* 59(4), 628-43.

Benkler, Y. (2006): *The Wealth of Networks*, New Haven: Yale University Press.

Berners-Lee, T., J. Hendler, and O. Lassila (2001): "The Semantic Web," *Scientific American*.

Bjorneborn, L., and P. Ingwersen (2004): "Toward a basic framework for webometrics," *Journal of the American Society for Information Science and Technology*, 55(14), 1216-1227.

Boyd, D. and N. Ellison (2008): "Social Network Sites: Definition, History, and Scholarship," *Journal of Computer-Mediated Communication*, 13(11), October 2007.

Chin, J. and P. Coveney (2004). Towards tractable toolkits for the grid: A plea for lightweight, usable middleware. UK e-Science Technical Report UKeS-2004-01. Available at: http://www.nesc.ac.uk/technical_papers/UKeS-2004- 01.pdf. Accessed: 3rd May, 2005.

Contractor, N., Poole, M.S., Srivastava, J. and D. Williams (2008): "Virtual Worlds: An Exploratorium for Theorizing and Modeling the Dynamics of Network and Group Behavior," presentation at Fourth International e-Social Science Conference, Manchester, 18-20 June.

Diani, M. (2001): "Social Movement Networks Virtual and Real," Information, Communication & Society, 3(4), 86-401.

Diani, M. (2003): "Networks and Social Movements: A Research Programme," in Social Movements and Networks: Relational Approaches to Collective Action, ed. by M. Diani, and D. McAdam. Oxford University Press, Oxford.

González-Bailó, S. (2007): "Mapping Civil Society on the Web: Networks, Alliances, and Informational Landscapes," DPhil Thesis, University of Oxford.

Jackson, M.H. (1997): "Assessing the structure of communication on the World Wide Web," *Journal of Computer-Mediated Communication*, 3(1). Retrieved September 19, 2008 from: http://jcmc.indiana.edu/vol3/issue1/jackson.html

Jackson, M. O. (2004): "A Survey of Models of Network Formation: Stability and Efficiency," in *Group Formation in Economics: Networks, Clubs and Coalitions*, ed. by G. Demange, and M. Wooders. Cambridge University Press, Cambridge UK.

Jackson, M. O. and B. W. Rogers (2005): "The Economics of Small Worlds," *Journal of the European Economic Association*, 3, 617-627.

Lusher, D and R. Ackland (2008): "A relational hyperlink analysis of an online social movement: Asylum seeker advocacy groups in Australia," submitted.

McPherson, M., Smith-Lovin, M. and J. M. Cook (2001): "Birds of a Feather: Homophily in Social Networks," *Annual Review of Sociology*, 27, 415-444.

Newman, M. E. J. (2002): "Assortative Mixing in Networks," *Phys. Rev. Lett.*, 89, 208701.

Park, H. W., and N. Jankowski (2008): "A hyperlink network analysis of citizen blogs in South Korean politics," *Javnost-The Public*. Vol. 15, No. 2, 57-74. A special issue on methodological issues in conducting online political communication research.

Rogers, R. and N. Marres. (2000): "Landscaping Climate Change: A Mapping Technique for Understanding Science and Technology Debates on the World Wide Web." Public Understanding of Science 9(2), 141–163.

Shadbolt, N., T. Berners-Lee, and W. Hall (2006): "The Semantic Web Revisited," *IEEE Intelligent Systems*, 21(3), 96-101.

Shumate, M., and L. Dewitt (2008): "The North/South Divide in NGO Hyperlink Networks," *Journal of Computer-Mediated Communication*, 13, 405-428.

Tapscott, D. and A. Williams (2007): *Wikinomics: How Mass Collaboration Changes Everything*, Atlantic Books.

Thelwall, M. (2002). Conceptualizing documentation on the Web: An evaluation of different heuristic-based models for counting links between university Web sites. *Journal of the American Society of Information Science and Technology*, 53(12):995–1005.

Thelwall, M. (2004). Methods for reporting on the targets of links from national systems of university Web sites. *Information Processing & Management*, 40(1):125–144.

Thelwall, M., Vaughan, L. and L. Bjorneborn. (2005): "Webometrics," *Annual Review of Information Science and Technology*, 39: 81-135.

Thompson, S. (2006): "Adaptive Web Sampling," *Biometrics*, 62, 1224-1234.

von Hippel, E. (2005): *Democratizing Innovation*, Creative Commons. License

Wasserman, S. and K. Faust. (1994): *Social network analysis*, Cambridge University Press, Cambridge UK.

Welser, H., Gleave, E., Fisher, D. and M. Smith. (2007): "Visualizing the Signatures of Social Roles in Online Discussion Groups," *Journal of Social Structure*, Vol. 8, No. 2.

Wimmer, A. and K. Lewis. (2008): “Below and Beyond Racial Homophily. ERG models of a friendship network based on Facebook.com,” Paper given at International Sunbelt Social Network Conference, St. Pete (Florida), January 22-27, 2008.

Yang, X, Wang, X and R. Allan. (2006): “JSR 168 and WSRP 1.0 - How Mature are Portal Standards?” Proc. WEBIST 2006 (WEBIST 2006), Setubal, Portugal, 11-13 Apr 2006.

[1] Note that some members of the social network analysis research community, as represented by the International Network for Social Network Analysis (INSNA, http://www.insna.org) and the SOCNET Usenet group, have expressed concern that the term "social networks" is increasingly being seen as being analogous to social media sites such as MySpace and Facebook. Boyd and Ellison (2008) recommend the use of the term "social network sites" for Facebook and similar sites, while Wikipedia uses the term "social network services" (http://en.wikipedia.org/wiki/Social_networking). In the present paper, we follow Wikipedia and use the term "social network services" to refer to the underlying technology (e.g. websites such as Facebook), while we use "online social networks" to refer to the social structures that are enabled by these technologies.

[2] Note that there is in fact active cross-over between the various approaches.

[3] This research as focused on large-scale networks more generally, with the Web as one such example.

[4] It should be noted, however, that there is growing interest in networks amongst economists and the work of Jackson (2004) and Jackson and Rogers (2005), which focuses on large-scale network such as the Web, is particularly relevant here.

[5] A listserv devoted to encouraging network scientists to study political phenomena, and political scientists to use the tools of network analysis, PN-L, was established only in July 2008.

[6] The Semantic Web has been defined as “the web of data with meaning in the sense that a computer program can learn enough about what it means to process it” (http://www.w3.org/People/Berners-Lee/Glossary.html#Semantic)

[7] SocSciBot is a long-established web crawler that is a popular data source for social science hyperlink analysis. See http://socscibot.wlv.ac.uk/.

[8] http://vlado.fmf.uni-lj.si/pub/networks/pajek/

[9] IssueCrawler is pioneering software enabling the collection and analysis of hyperlink data, and is popular in the humanities and social sciences (see, for example, Rogers and Marres 2000). See also http://www.issuecrawler.net/.

[10] See Lusher and Ackland (2008) for further discussion of the differences between SocSciBot, IssueCrawler and VOSON. See also http://voson.anu.edu.au/.

[11] http://code.google.com/apis/opensocial/. The original intention was to build a Facebook application, but OpenSocial appeared easier to work with and the potential for using it on different SNSs (it will work under Ning, Orkut, MySpace and a number of other SNSs) was attractive. The OpenSocial API is still under development and is not yet widely available to developers. There is evidently a lot of interest in OpenSocial in the Indian web software development community, and as of April 2008, only Orkut

users who have selected "Indian" as their nationality in their profile are able to develop and use Orkut OpenSocial applications (this is why the authors nationality is stated as Indian in Figure 1).

[12] http://www.gridsphere.org/

[13] http://www.jsviz.org/blog/

[14] Note that the map in Figure 1 contains dummy data since the author doesn't actually have any friends on Orkut.

[15] Note that these metrics should actually be calculated for a complete network, rather than an ego network.

[16] http://erzuli.ss.uci.edu/R.stuff/

[17] http://jung.sourceforge.net/

[18] http://www.myexperiment.org/

[19] Sampling is fundamental to empirical social science research but relatively foreign to those whose research involves searching over collections, since one would never take a sample of a collection (all members in the collection need to be locatable).