Abstract

Political communication research is increasingly concerned with the study of political life in online environments. Some recent investigations make use of Internet-based tools for the research process: for project management, for data collection and analysis, and for the preparation and publication of findings. In these respects, political communication research reflects methodological transformations underway across the social sciences, often known as e-Science and e-Research. This article explores aspects of that transformation through examination of a range of studies concerned with political discourse, political participation, and election campaigns in which the online environment is accentuated. We reflect on four study features: project management, research designs, sampling, and data visualisation. In a series of illustrative studies, we consider challenges in undertaking political communication research in network environments utilising Internet-based tools. Finally, we introduce the contributions to this journal theme issue, placing the articles within an overall framework of concern regarding Internet-based political communication research.
Introduction

One of the initially expressed promises of the Internet, especially of the Web, relates to the potential such technologies might have for political reinvigoration. The history of the Internet is rich in utopian expression, with early commentators extolling the political virtues or vices, frequently without empirical substantiation for the claims and predictions made (e.g., Davis & Owen 1998; Rash 1997; Margolis & Resnick 2000). With time, empirically-oriented scholars began to examine the claims more closely and since those early days considerable research has been published that explores facets of the claims: the nature of and contribution to politically-oriented discourse, the specific contributions of the Internet to information provision and political action during election campaigns, and the more general impact of the Internet on political awareness and action (e.g., De Vrees & Schmitt 2007; Gibson, Nixon & Ward 2003; Norris 2000; Shane 2004).

These developments are reflected in a body of literature that has been reviewed and summarised elsewhere (e.g., Lusoli 2005). This literature and the associated reviews, however, mainly consider the substantive issues of political communication in online environments; few contributions address methodological concerns related to conducting political communication research in such arenas. This task is the central objective of this theme issue of Javnost – The Public, and in this introductory article we lay the groundwork for this exploration.

We begin by delineating central terms and areas of concern: empirically-based political communication research, an overview of methodological issues in the social science investigations that are important to Internet-oriented studies, and consideration of relatively new methodological developments known as e-Science and e-Research. This “setting of the stage” is followed by a range of illustrations where common methodological issues are involved: in the general matter of project management and researcher collaboration; in the research design of studies and use of mixed method approaches; in the sampling procedures and relation of findings to larger groups; and in the visualisation of findings, particularly through social network analysis. Although many other themes could be addressed, we feel the issues related to these themes are especially relevant for consideration in conducting political communication research in online environments.

The next section of the article provides a summary of the points made and suggests possibilities for methodological innovation regarding Internet-oriented political communication research. This section leans on previously formulated summaries of methodological innovations composed for other purposes (Jankowski & Van Selm 2005). Finally, we briefly present the five articles prepared for this theme issue, suggesting points of importance, points for comparison, and areas meriting further methodological exploration.

Parameters of Concern

Although it would be excessive to explore basic terms such as politics and political communication research in the context of this article, it is appropriate to acknowledge that we subscribe to a broad vision of what constitutes political communication research, extending beyond the traditional foci of political scientists on electoral campaigns and formal political systems. “Politics is everywhere” Brants
(2002, 187) observed, and we would add that political concern encompasses the breadth of everyday life along with political system-determined moments such as election campaigns and voting.

Internet research is frequently seen as investigations that have their focus situated on the Internet, (e.g., studies of political expression and action on Internet discussion lists) and investigations that make use of network environments for data collection and analysis (e.g., hyperlink analysis, email surveys). This distinction is heuristic: studies concerned with Internet-based phenomena are often conducted with the aid of instrumentation facilitated or made possible by the Internet (e.g., a study of political discourse on Weblogs through a Web-based survey of bloggers). Although some disciplines in the social sciences address these two forms of Internet research, particularly the methodological issues, political science seems to be a lagging in this respect. For example, a relatively recent methodology textbook, *Research Methods in Politics* (Burnham, Gilland, Grand & Layton-Henry 2004), devotes a single chapter to some of the issues (e.g., representativeness of respondents, reliability of resources, ephemeral nature of materials) and much of this chapter is reserved for discussion of data available via the Internet (e.g., online databases). This degree of attention contrasts strikingly to that reflected in a range of book-length treatments situated within other social sciences (e.g., Jones 1999; Hine 2000, 2005; Johns, Chen & Hall 2004).

The above-mentioned attention to forms of Internet-based research is a special case of what is frequently labelled e-Science or e-Research. The first term, e-Science, has its grounding in the natural sciences and engineering, and primarily refers to utilisation of network-based tools for enhancing distant collaboration and facilitating access to and analysis of very large sets of data using super-computers linked together in what is known as a "Grid" (see, e.g., Buyya & Venugopal 2005). Several initiatives are underway to adapt e-Science practices to the humanities and social sciences, and the more general term of e-Research reflects these initiatives (e.g., Genoni, Merrick & Wilson 2008 forthcoming). Basically, this term refers to a set of six components: international distant collaboration of investigators, use of high-speed computers, visualisation of data, development of Internet-based tools and research procedures, construction of virtual organisational structures, and Web-based distribution and publication (Jankowski, 2007).

**Illustrations of Methodological Issues**

Space prevents attention to all of the above-mentioned components within the confines of this article. We focus on four components with relevance for political communication research conducted in online settings: project management and collaboration, research design, selection of cases, and data visualisation.

**Project Management & Collaboration**

A study illustrating some of the challenges associated with project management and collaboration at a distant is the Internet & Election Project, launched in 2003 and culminating in publication of an edited volume on the work (Kluver, Jankowski, Foot & Schneider 2007). Research teams involving 30 investigators examined how the Web was employed by a wide range of political actors during national election campaigns in 19 countries across Europe, Asia and North America in the period
2004-05. The global dispersion of the research teams made project management and collaboration particularly challenging. Few moments could be created for face-to-face meetings, the most important being a three-day workshop planned early in the project to inform participants about objectives and train them in use of the coding instrument for examining campaign-related Web sites. In addition, some participants were able to meet during international conferences. Other project communication transpired via telephone conferences, email, and instant messaging — all of which reflected both the limitations of such mediated forms of communication as well as the possibility of such distant research collaboration. The value of that duality stated, it is important to stress that this project was only possible thanks to these distant communication “tools,” involving thousands of email messages and hours of instant messaging, particularly during the intense periods of data collection just prior to the elections. As demonstrated in studies of distant collaboration among researchers, a basis of trust developed during periods of personal encounters largely determines the degree of success of such projects in which participants are required to collaborate for long durations in different locations (see, e.g., Olson, Zimmerman & Bos 2008 forthcoming). In the Internet & Election Project that basis had been developing prior to the project among the coordinators and a small core number of participants, and was increased and nurtured during the course of the undertaking through a series of face-to-face meetings.

Research Design

A large array of considerations take place during the design of a research project: whether to situate a project in a single or multiple sites, whether to employ one or a multiple of data collection methods, whether to construct a “one-shot” study within a specified time frame or to prepare a longitudinal study, whether to restrict the approach to an exploratory investigation or to test hypotheses related to a theoretical notion, whether to develop an interpretative and primarily qualitative study or one that emphasises quantitative measure. There are other design considerations (see, e.g., Turnšek & Jankowski 2008), but this list suggests the range of issues involved. One of the most common — and problematic — issues from the list involves intertwining qualitative with quantitative methods of study. Howard (2002) addresses this issue with regard to ethnography and social network analysis, and proposes a blend that he terms “network ethnography.” This mixture is based on a somewhat traditional conception of ethnography involving long-term immersion in a culture as opposed to the relatively superficial forms of “virtual ethnography” that often utilise limited sources of data (e.g., email and chat room exchanges) and select cases without adequate consideration. Such in-depth, rich-data ethnography, Howard suggests, can have a synergistic effect when combined with social network analysis and is “especially useful for studying communication in modern organisations over new media” Howard (2002, 552). Further, such combination can create a balance between “macro-structure” and “micro-agency” concerns, according to (Howard 2002, 570).

Similar multiple method designs are frequently proposed and may involve combining any number of data collection methods such as discourse analysis with quantitative content analysis and survey research with expert interviews. Often, these integrations of data collection are undertaken to achieve “triangulation” of
sources that is felt to provide greater validity of findings. Such notions of triangulation, however, are all too frequently based on the assumption, not necessarily true, that agreement in research findings from different sources provides additional assurance that the findings are valid; it is just as plausible that findings from both data sets are biased in the same direction. Although Howard does not discuss triangulation and its problems, he does correctly note that findings become richer and more reflective of an object of study when based on multiple methods; the fusion he proposes between ethnography and social network analysis is particularly suitable with various Web 2.0 platforms such as social networking sites, blogs, and video-sharing sites.

Selecting Cases

Of critical importance in any study is the procedure by which objects are selected. In survey research the procedure is known as sampling and similar procedures are frequently applied to other forms of data such as content analysis (Van Selm & Jankowski 2006). Most sampling procedures relate to probabilistic sampling whereby each object of study (e.g., persons, documents) has an equal chance for selection. The basic and generally preferred form for such selection is simple random sampling, which allows a researcher to estimate with a statistically determined degree of certainty probable characteristics of the population from which the sample was drawn.

In contrast, non-probabilistic sampling deviates from the basic principle of equal chance and, as a consequence, generates uncertainty as to whether the results found reflect those of the population from which the sample was drawn. There are many forms of non-probabilistic sampling procedures: theoretical sampling, extreme or deviant case selection, snowball or viral sampling. Although these procedures may have value in particular cases, they are plagued by the potential bias built into self-selection of cases. Researchers, for example, that place invitations to complete an online questionnaire on public Web sites or discussion lists are inviting response weighted towards persons with personal reasons to take part in the studies, and when such researchers subsequently report the findings as if they are reflective of a more general population a basic error in sampling practice is being made (see, e.g., Boogers & Voerman 2003).

It may be valuable to review some of the basic features of sampling as applied to selection of Web sites for conducting a content analysis. The terms and procedures are directly derived from survey research and content analysis performed in more traditional, offline venues. First, determination of the sampling frame (from where to sample in order to represent the phenomenon under study) is important, along with explication of the properties of the sample units (determining which parts of a Web site are to be investigated). As for sampling frame, the size and dynamic, ever-changing character of the Web complicates drawing a random sample of objects for study by conventional scientific procedures. In now classic literature in the realm of Internet research, McMillan (1999) considered scientific sampling one of the most difficult aspects of content analysis of Web-based material, given that the number of Web sites is not constant and that the available indexes of sites are incomplete and overlapping. Weare & Lin (2000, 280) posed several alternatives for drawing samples from the Web, using various sampling frames – domain names,
search engines, and sites providing overviews – in addition to direct random sampling procedures.

When domain names are used to determine the sampling frame it is important to be aware of the hierarchical structure of these names, the Domain Name System (DNS), and the decentralised manner in which the system is administered. In their study, Weare and Lin (2000) distinguished various levels of domains: top-level domains (e.g., org, edu, int, eu); second-level domains (e.g., usc.edu, yahoo.com, and specific groups of sites such as Californian sites (ca.us), commercial sites located in a specific region (e.g., uk.com); and third-level domains designed to further specify the site (e.g., annenberg.usc.edu). As a consequence of decentralised governance of Web site URLs, the hierarchical structure of the domains varies. In spite of this, Weare and Lin regard a list of domain names a useful sampling frame, mainly because of the hierarchical nature and completeness of the DNS, especially when the research question maps neatly into a domain-based category. A limitation of employing a DNS-based sampling frame, however, is that the frame will not include Web pages nested within a directory structure of a host computer. Therefore, this sampling frame is better suited for broad questions of Web content, and cannot be used for research questions that focus on lower level (such as sites of department within universities or divisions within a corporation). Other limitations mentioned include bias from sampling sites within a specific domain, and the chaotic manner in which the Internet has developed, which can make reliance on a specific domain problematic.

A second relatively popular method for constructing a sampling frame involves using search engines. Weare and Lin mention several advantages to this approach, such as cost and extending the sampling frame beyond first-level domains. A serious disadvantage to this method is that the Web is not consistently and universally catalogued, which can lead to samples skewed to the more heavily trafficked parts of the Web. This may not be problematic in studies of the populations of sites that an average Web surfer would locate, but in studies addressing other research populations, such as sites targeted at closed or exclusive communities, this approach to selecting cases may be problematic. Although use of meta-search engines such as MetaCrawler may not be able to eliminate all possible bias, their use is recommended inasmuch as they generate lists based on several search engines, and, thus, lists of sites that are more comprehensive (Weare & Lin 2000, 279).

Once a sampling frame has been established, the next step, particularly for content analysis, is determination of sampling units. Selection of these units depends mainly on the research question to be addressed. Defining the categories “context units,” “recording units,” and “units of analysis” requires taking decisions regarding the focus of study, such as an entire Web site or a smaller unit such as a single page on a site (McMillan 1999). Whereas in survey research the sampling units (respondents) are at the same time the recording units (respondents filling out the questionnaire) and the units of analysis (characteristics of respondents as reflected in tabular presentation of results), in content analysis this is seldom the case. Generally, in content analysis the sampling units contain larger entities (e.g., a newspaper edition, a television commercial) than the recording units, that is, those parts of the material to which coders can reliably pose their questions (e.g., headlines of news articles, newspaper photographs, or main characters featuring
in a television commercial). Although there is considerable experience in defining (the relationships between) sampling and recording units in the content analysis of traditional media materials, this experience is much less extensive for analysing Web sites.

Many content analytical studies conducted on Web-based material employ the entire Web site as recording unit. Weare and Lin (2000, 282), however, suggest that “relying on coders to evaluate an entire site as a whole is unrealistically demanding.” Another option is to consider the Web page as the recording unit, although this is a rather aggregated recording unit, comparable to an entire newspaper article. Li (1998) considers as unit of analysis a single day of publication of an electronic newspaper, and the recording units as the front page and news articles.

In a study of the content of the Web sites of Dutch political parties Van Selm, Jankowski and Tsaliki (2002) consider the pages available on the first three “layers” of the Web sites of three Dutch political parties as recording units. The research material consisted of the first three layers of pages belonging to the Web sites of the parties with URLs: www.groenlinks.nl, www.sp.nl, www.cda.nl. For example, the first layer consists of the welcome or starting page of the web site. The second layer refers to those pages that can be opened by means of a button placed on the welcome page. Hence, the second layer represents pages that are encountered when visitors choose to click on a hyperlink available on the welcome page. Pages on this layer typically share the URL “stem” of the welcome page. The third layer includes those pages that can be opened by means of a button placed on one of pages located at the second layer. The authors decided to analyse only three layers of each Web site inasmuch as the number of pages grew almost exponentially with every layer. Examining Web sites in this hierarchical manner could be criticised since visitors are able to reach pages belonging to the site in many ways (e.g., via hyperlinks found on other Web sites). In an interview with a representative from one of the political party sites, however, the respondent estimated that more than 75% of site visitors initiated visits via the party homepage.

Visualising Data

The visualisation of data is perhaps as old and venerable as the first efforts to present data in tabular form. Such visualisation can be elementary, as in the frequency table of individual survey variables and two-by-two cross table correlations, and they can be complex, as in the tabular presentation of multivariate findings or log-linear analysis. The underlying objective of these visualisations is to reduce the complexity of data in order to increase understanding of the relationships among variables. Most forms of such data visualisation consist of numbers and statistical measures in the cells of two-dimensional tables; an occasional and unusual form of visualisation in conventional social science research is illustration in the form of an histogram or pie-shaped figure reflecting the relative size of the parts to the whole.

These conventional visual presentations of data are undergoing rapid change as social network analysis emerges as the preferred approach to reflect and understand relationships in a networked environment. Relationships among bloggers can be presented (see the Park and Jankowski article in this issue), as can the connections found in YouTube videos and discussion lists. These last mentioned venues, often
collectively known as Web 2.0 platforms, are a particularly rich terrain for social network analysis and the accompanying visualisations showing different size and coloured circles (nodes) connected by lines of differing thickness. The visualisation of Web 2.0 platforms related to political communication concerns are only now beginning to appear in conference papers and recent journal publications (e.g., McKelvey 2008; Devereaux 2008; Langlois 2008; Black & Welser 2008). Although there is frequently an aesthetic attraction of these visualisations and an implicit suggestion that the images provide clarity to the relationships under study, there is all too often a failure to provide interpretation to the images under the assumption that they “speak for themselves.”

Although the present exploration of Web 2.0 platforms related to politics may be in its early days, use of social network analysis has a rich and long history, particularly within divisions of sociology (see, e.g., Wasserman & Faust 1994). Some applications can be found in the domain of political communication research (e.g., Jankowski & Van Selm 2000) and codification of the procedures is also developing for this area. Felling and Van Selm (2006), for example, explore social network analysis to the field of Internet communication and focus on mapping ego-centred and entire networks. Analyses of ego-centred networks can provide insight into the extent to which network members are tied to each other, and in which actors have access to sources of information. Analyses of entire networks map which members occupy central or peripheral positions in a network. In social network research the unit of analysis is the relationship between actors of which characteristics are assessed, such as its direction (who is donating respectively reacting to information) or strength of the relationship (e.g., contact frequency). Empirical literature about network analysis in the domain of CMC can be structured along at least two general research questions. The first type, often based on log data, aims at analysing characteristics of the electronic networks itself, such as an electronic political debate (e.g., Hagemann 2002). The second type of study examines the relative role of Internet deliberation compared to communication through other media in building and maintaining a (temporary) social network.

This is not the place to elaborate in detail on a particularly complex form of analysis, but two general reservations merit mention. First, there can be the temptation to elevate the aesthetics of the visualisation above and beyond analytic interpretation related to the research concern. When this happens the presentation of the project may be more reflective of artistic values than scholarly argument. Perhaps the most challenging development in this area is the visual presentation of network change across time. Researchers at Ryerson University in Toronto are exploring use of tools such as Google Charts for the longitudinal change in the number and focus of comments related to YouTube videos associated with candidates in the Canadian elections (Devereaux 2008). Although such dynamic visualisation presents the temporal dimension of the relationship in an intuitively appealing manner, it is unclear how such illustrations translate into social science interpretation. Second, visualisations of social networks can be particularly valuable in the exploration of relations, but not in uncovering the substance of the relations. For this concern, other sources of data are required, as illustrated in a study on the nature of deliberation involved in preparing an article for Wikipedia in which social network analysis is combined with discourse analysis (Black & Welser 2008).
Conclusion

In this introduction we have tried to indicate the importance and challenge of addressing methodological issues related to investigating political communication in online arenas. Many of the challenges have been on the agenda of social scientists for decades and concern such matters as designing studies and selecting cases. Such generic methodological topics often require special attention in online contexts. For example, as we have shown, sampling becomes problematic when there is little assurance that the object of study – say, a Web site – is continuously in flux or when the search engine used to identify sites produces different results at different times of consultation. Conducting research in such a dynamic environment brings into question assumptions many of us nurture: that the world and social life are sufficiently constant to allow replicable exploration and scientific prediction. That difficulty stated, the challenge remains to address these features in a manner that reduces uncertainty and unclarity in the methods employed for doing online research.

This has been the challenge, in fact, of a growing genre of literature about online research. Many reviews of this literature have been prepared (e.g., Jankowski & Van Selm 2005), but because of the continual development of online communicative tools and their application in the arena of politics, such reviews are in need of periodic revision, and this theme issue is designed to contribute to such updating. While change is integral to online environments and to the methods for investigating communication in those environments, there remain perennial methodological concerns: how to conduct an appropriate study design, how to select cases relevant to the focus of the study; how to manage a cross-national or multi-site project involving researchers situated in disparate locations; how to visualise networked data often of a dynamic and multimedia nature; how to employ the tools and procedures associated with e-Research including; how to ensure preparation, publication and dissemination of research findings in outlets suitable for and able to reflect the multimedia richness of much online communication.

The five articles in this theme issue consider a number of the methodological matters important to conducting political communication research in online environments. Although far from all considerations are present, the collection provides contextual detail for a *capita selecta* of concerns. The detail is both rich and specific for concrete research projects, which is the strength of the collection. Todd Graham leads the issue with “Needles in a Haystack: A New Approach for Identifying and Assessing Political Talk in Nonpolitical Discussion Forums.” He argues persuasively that political discourse can be found in the most common of situations, including online discussion forums related to reality TV programs such as Wife Swap and Big Brother. Finding and analysing political discourse in such forums, however, requires special search procedures and a form of analysis sensitive to the informality and subtlety of the communicative exchanges.

Maurice Vergeer and Liesbeth Hermans present a mixed method approach for analysing political discussion in “Analysing Online Political Discussions: Methodological Considerations.” The authors elaborate on the basic features and potentials of social network analysis for understanding political exchange in Web environment and added value in combining such an approach with content analysis across time.
The focus of their study is a discussion list devoted to political topics, for which they examine a sample of contributions to threads with this approach and reflect on the suitability of this mixed method design.

Han Woo Park and Nicholas Jankowski also perform a social network analysis, but of election-related blog posts in “A Hyperlink Network Analysis of Citizen Blogs in South Korean Politics.” They are especially interested in the relational features of such politically-oriented posts: in how citizens and politicians relate to each other in such public forums for political discourse. The analysis concentrates on the in and out links of the posts and, combined with examination of the topics of the blogs, provides insight into the differences in the directions of political discourse among those with political power and members of the electorate.

Tamara Witschge, like other contributors to this issue, develops a multiple method design to explore political exchange: “Examining Online Public Discourse in Context: A Mixed Method Approach.” Her design includes a form of discourse analysis combined with an online survey held under participants in the discussion forum studied. She incorporates attention to the context of the exchanges and, together with a form of critical discourse analysis, unravels the nature of exchange on a both salient and sensitive topic to members of the group.

Wu Mei concentrates on the challenges involved in “Measuring Political Debate on the Chinese Internet Forum.” The author also employs a multiple method approach, combining both qualitative and quantitative forms of content analysis. With a two-stage design, meta-data from the discussion forum are used to focus analysis during the second stage on qualitative characteristics of the political posts. The design and tools especially developed for the study may have relevance and application to other online discussion forums in other contexts and countries.

Acknowledgements

This theme issue was initiated during and benefited by discussions held at meetings of COST Action A30, “East Meets West: Setting a new Central Eastern European Media Research Agenda.” The first author, a member of this COST Action, wishes to express appreciation for the opportunity to present aspects of this research during meetings of the COST Action.

Notes:

1. For an extended, critical discussion of triangulation, see Schroder, Drotner, Klein & Murray (2003).

2. Such fusion of methods is under consideration in a study being launched around the uptake of Web 2.0 platforms during the 2009 European Parliament election campaign; see Jankowski & Voerman (2008).

3. Illustrative of such an assumption of images “speaking for themselves” is an article with six figures derived from social network analysis, all of which are “interpreted” in a single sentence; see Rogers (2008).

4. In a training workshop on the use of software tools for network analysis (Issue Crawler) that the first author of this article attended, much attention was given to the appearance of the visualizations; computer designers were given the task to manipulate the images in an aesthetically pleasing style prior to public presentation of the material.
References:


