ANALYSING ONLINE POLITICAL DISCUSSIONS: METHODOLOGICAL CONSIDERATIONS

MAURICE VERGEER

LIESBETH HERMANS

Abstract

Online political discussions are thought to lead to more political engagement and empowerment of peripheral groups in society and thereby contributing to deliberative citizenship. Because people have increased opportunities to voice their political opinions and publish these for a potentially large audience to read, the involved level of interactivity can mobilise people who would otherwise not have been in political life. Since Web 2.0 applications (i.e. blogs, social networking sites) have become popular, online discussions have taken a great flight on the web. This article discusses the advantages and difficulties of studying online discussions applying a mixed method approach of content analysis, social network analysis and longitudinal analysis. The additional value of using a combination of research methods simultaneously is that it does justice to the complex object of study because a more in-depth and triangulated measurement of political communications can be established. The methodological implications will be illustrated on data from the online political discussion group, nl.politiek, one of the most active discussion groups during the Dutch national elections in 2006.

Maurice Vergeer is Associate Professor in the Department of Communication, Radboud University Nijmegen; e-mail: m.vergeer@maw.ru.nl.

Liesbeth Hermans is Associate Professor in the Department of Communication, Radboud University Nijmegen; e-mail: l.hermans@maw.ru.nl.

37 javnost–the public Vol.15 (2008), No. 2, pp. 37-56

Introduction

Online political discussion forums, like newsgroups, chat rooms, and blogs are relatively new interactive communication innovations that contribute to an electronic democracy (Davis 2005; cf. Shane 2004). Online communication has transformed the public sphere in the sense that people (i.e. non-professionals) have increased opportunities to disseminate their thoughts unmoderated over a wider geographical area than prior to the internet.

There are many ways to study online discussion groups. Most studies focus on the content of the discussion in messages (cf. Hill & Hughes 1997; Papacharissi 2004), while others focus on the visualisation of the structure of the large data sets on discussions (Sack 2001; Turner et al 2005). In this article, we discuss ways to analyse political online discussions using mixed methods. Using a combination of content analysis with a social network approach, we analyse data within a longitudinal perspective. The additional value of the combination of multiple research designs and measurements is that it allows for triangulation of results, meaning that the same object under study is studied from multiple angles. These combined research strategies produce a more valid measurement and therefore a better understanding of the complex, multidimensional (political) content of these new communication forms and therefore can contribute to a fuller understanding of how discussion groups interact with the political processes in a democracy.

The methodological implications will be illustrated on data from the Dutch online political discussion group, nl.politiek, one of the most active discussion groups during the Dutch national elections of 2006. The content of electronic discussions consists on typed messages in which individuals share their opinions and comments on political issues. Usenet, facilitating citizens' participation in the political process, is often discussed in the context of the deliberative democracy (cf. Davis 2005; Shane 2004) The non-hierarchical and decentralised character of the discussion groups enables communication between people who would otherwise not have been active in the political discussion.

Political Communication and the Internet

Although the use of the internet by the general public only became substantial during the past decade, considerable study has already taken place on the potentials and limitations of online political communication within a diversity of platforms. See, for example, studies of websites maintained by a range of political actors such as political parties (Van Os, Jankowski & Vergeer 2007), politicians and non-governmental organisations (Shane 2004), hyperlink analyses of websites and blogs (Park 2003; Tremayne et al 2006), and studies of the characteristics of the users of online political communication (Davis 2005; Hill & Hughes 1997; Norris 2001).

The internet appears to have polarised political observers into optimists and sceptics (cf. Tedesco 2004; Davis 2005). The so called e-optimists view the internet as more than obtaining information, "but as revolutionizing the character of democratic society by transcending limitations of time, space and access and interactive and deliberative citizenship, not hindered by the elite character of traditional mass media" (Brants 2005, 143). The potential interactivity will draw an increased mobilisation of people into the political discussion. At the other end of the spectrum

there are e-pessimists who question the potential power ascribed to the internet in activating the politically uninterested. These pessimists worry because digital power can create a new digital divide between those who do and those who do not have access to the communication resources. Although access to the internet has increased, people still need special skills to participate in online communication (Norris 2001). A more realistic view as compared to the e-pessimists and e-optimists is offered by those who conclude that it is still too early to evaluate the impact of online communication on the political arena. Although little has changed in the political process, new possibilities created by the internet may lead to empowerment of peripheral groups (Muhlberger 2004).

Further empirical research is necessary to support the argument used in this scholarly debate (e.g., Dahlgren 2002, 2005; Savigny 2002).

Online Political Discussion, Public Opinion and Political Systems

Online political discussion forums such as newsgroups were one of the first formats for political communication within an electronic network environment. Because electronic discussions enable people to express their political views to a large potential audience, online discussions by citizens may provide new opportunities for political participation. People who felt ignored before in the traditional political arena could participate in the alternative online arena on public issues (Davis 2005). The increasing usage of online discussions for persuasion and mobilisation suggest that electronic forums are important in the formation and expression of public opinion (Davis 2005). However, Davis (1999, 165; 2005, 117) argues that, in the United States, those who post messages on the online discussions are a small and atypical minority that cannot be seen as representative of the population. He compared online discussants' demographic background, political attitudes and behaviour, with the general public, and then compared if the issue interest between them differed. Because he did not include content analyses of the online discussion, it is still difficult to judge whether opinions expressed in online forums represent public opinion.

Furthermore, research on political discussion groups has been conducted mainly around sites focusing on events in the United States, but this political-media system is not necessarily reflective of other systems, particularly of the political system in the Netherlands. For example, the political system in the United States is based on a two-party majoritarian system and a media system characterised by "a neutral, commercial press and information oriented journalism." These characteristics differ considerably from those of the Dutch system, which has a multi-party system based on consensus and which has a media system with "a long dominance of a party linked public broadcasting system" (Brants & Van Praag 2006, 27; Hallin & Mancini 2004).

It is important to take the different political contexts into consideration when studying and interpreting results of the research on political online discussion groups. In the U.S., many discussion groups are principally ideologically oriented (Davis 2005; Hill & Hughes 1997), whereas in the Netherlands, only a few general political discussion groups exist without a specific political ideology. The content of messages within discussion groups will be different because with more homogeneous ideological participants content will indicate the prevailing ideological

consensus of that group (Hill & Hughes 1997). In the non-ideological discussion groups, there probably will be a more heterogeneous ideological contribution. In the section of this article on comparative analysis we address this topic.

Internet and Online Discussions

Online discussions can take many forms on the internet. One of the first applications was Usenet, a distributed system of servers hosting a large number of discussion groups (cf. Schneider 1997; Smith 1999). Usenet consists of predominantly text based discussions, organised by hierarchically organised themes. Usenet discussions groups can be accessed through newsreaders, and increasingly through a web interface (Davis 2005). In 2001, Google (2001) acquired Deja.com a large provider of access to Usenet. Subsequently, Google allowed free access, available through their search page, to all discussion groups since 1995.

People can start a discussion (i.e. a thread) by posting an article or reply to an earlier posted article. This kind of discussion is termed a many-to-many form of asynchronous communication. Discussion groups as such can be seen as thematically structured virtual communities. Those communities differ from "real world" communities because they have no geographical boundaries (cf. Jankowski 2002). Instead, these communities solely exist based on a shared interest in a specific, narrowly defined topic. People share their interests but would have never communicated with each other without Usenet. Aside from a web interface to access Usenet discussion groups, other forms of online discussions are nowadays available on the web. For instance, web logging or blogging (cf. Tremayne 2007a) on a wide range of topics has become popular (cf. www.technorati.com) as well as to express personal opinions on news articles in online newspapers.

These online discussion groups as virtual communities can be viewed as an aggregate of individuals and at the same time as individuals that are interconnected. An important criticism on this approach is that the individual is viewed isolated from its social structure. A paradigm and a research design that addresses this criticism is social network analysis. Predominantly developed within the sociological discipline, it was Rogers (1995) who translated the approach to the field of communication in his studies on the diffusion of innovations. The paradigm is well formulated by Rogers and Kincaid in the following two citations:

> *In real life, natural settings, communication can be understood better if it is not broken up into a sequence of source-message-channel-receiver acts, but rather examined as complete cycles of communication in which two or more participants mutually share information with one another in order to achieve some common purpose, like mutual understanding and/or collective action* (Rogers & Kincaid 1981, 31).

And:

> *In the study of human communication, we feel that emphasis should be placed upon information exchange relationships, rather than on individuals as the unit of analysis* (Rogers & Kincaid 1981, 32).

To date, network analysis plays a minor role in communication science. This is striking because the adoption of the internet on a large scale has opened up new possibilities for communication network analysis (Monge & Contractor 2003).

Methodological Considerations

Political online discussions can be analysed in several ways. The most predominant way is studying general public opinion in online discussions, such as describing what issues and political actors (e.g. parties and politicians) come forward and the evaluation of these issues and actors. The method for this approach usually involves a form of content analysis.

A second approach looks at the community aspect of the discussion group. A discussion group can be viewed as a community sharing opinions on a narrowly defined topic; in this case politics (cf. Jankowski 2002). However, within this community, people may have conflicting interests in varying degrees. These conflicting interests may follow, for example, the dimensions of left wing and right wing (e.g., liberal versus socialist) or religious and secular. As such, the discussion group is a platform for voicing conflicts of interests. To take the community approach a step further, discussion groups can be approached from a social network perspective. Wellman (2001) argues that computer-mediated communication networks link dispersed people with each other, resulting in social networks. By looking at the interconnectedness of newsgroup participants allows for community structure analysis.

In this study we will focus on the content analysis of online discussions regarding politics and of the communication networks between authors contributing to these discussions.

Organisation of a Discussion Group: Units of Analysis

Discussions in online groups have a hierarchical, sequentially and nested structure. A discussion group consists of multiple separate discussions. Each discussion consists of at least two messages, produced by at least two actors. Within each message, the identity of each actor can be more or less identified through supplied name and/or e-mail address in the message header and a signature in the message body. Discussion threads are hierarchically organised, meaning that the first message of a discussion is at the top (although the default sorting order is up to the user of news reader software such as Outlook Express, Newsbin Pro or Agent).

When an actor adds a new message to the discussion, this is embedded in the already existing discussion. Depending on the length of the thread, the actor can read the discussion and can take into account everything mentioned earlier. However, what people read is not logged on newsgroup servers, and therefore not measurable unobtrusively. To reconstruct discussions analytically, one can use so called header data, technical data contained in the messages. This header information contains information on who (e-mail address and/or name), replies to what other messages (reference-ID), from which computer (IP-address), and when (date and time).

Content Analysis

Research employing quantitative content analysis has a long tradition in studying and analysing media messages (Rife, Lacy & Fico 2005). Political media messages are produced and formatted by professional communicators, such as journalists and campaign managers. The media content is often produced on a routinely basis, moderated and edited by an editorial board before being published. The political content, such as political advertising, is produced on strategic basis by politicians

or spin doctors. This leads to content that is well thought through, clearly formulated and well substantiated. The content in discussion groups differs substantially because it is produced by non-professionals and is therefore more informal. For instance, non-professionals may act and react in an unrestrained and more straight forward. Furthermore, they will probably not be as capable of expressing their thoughts in written text. This leads to discussions that are, to some degree, more similar to conversations (Drew & Heritage 2006).

Newsgroups differ from each other in the nature of the discussions. Apart from the topic of discussions, the manner in which these discussions are held may range from on-topic and polite, to completely off-topic and offensive. Whether or not contributions are offensive has to do with the degree of implicit and explicit norms in the group (cf. Burnett & Bonnici 2003), resulting in varying degrees of social control. Often explicit social norms (cf. a charter or FAQ) on how to behave are unavailable. This means that unsavoury behaviour such as flaming, insults and discriminatory remarks are unpunished. Even if a charter is available, people still may choose not to conform to it. Furthermore, because people do not know each other personally, inhibitions to insult others are less present than when people do know each other personally. This is aggravated by the lack of visual cues aiding the interpretation of messages (cf. Goffman 1959). So-called emoticons may help the receiver to interpret online messages as they were intended. A final characteristic of online discussions is their asynchronous nature, meaning that there often is a time gap between an initial contribution and a reply. Due to the asynchronous nature of these discussions, quick corrections or swift punishment on abusive content is difficult. This hampers the quality of the discussion. All in all, this results in political content in online discussions less structured and more emotional than professionally produced content.

The specific nature of the online discussions has consequences for the performed content analysis. Whereas automated content analysis is best applicable when one is studying the more formal language used in professionally produced content, manually coded content analysis is preferred for analysis of everyday online discussions. The measurement instrument, depending on the research questions should reflect these peculiarities. Some of these peculiarities are spelling errors and the use of nick names that may be offensive and/or degrading.

Network Analysis

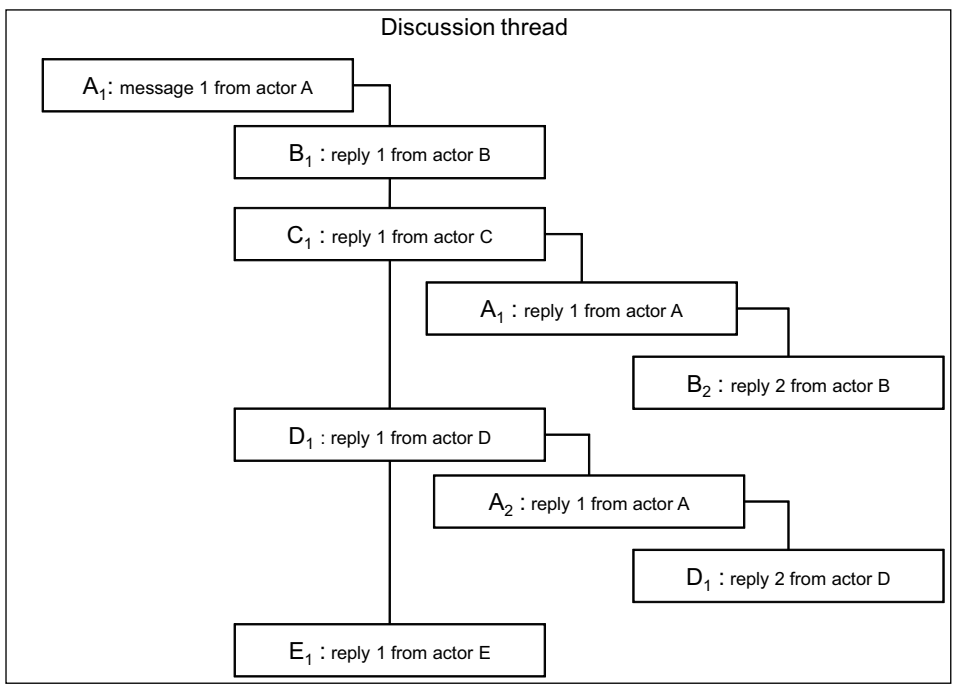
Network theory is an approach in social sciences developed as a critique on the atomistic approaches such as the survey designs where individuals are analysed isolated from their social context and the origin of variation is at the individual level, while community and society are considered beyond the scope (Rogers 1995). Social network analysis focuses on the social structure of one or more groups and on individuals with such groups.

Although network analysis is presented here as a method of analysis or even a methodology, a clear distinction must be made in "social networks on the net" and "social networking on the net." Web sites such as My Space, Facebook, Hyves and many more facilitate social networking as an activity in itself (cf. Boyd & Ellison 2007). While social networks on the net are the results of all sorts of communicative actions, social networking is a purposive activity in itself.

Units of Analysis

The unit of analysis in network analysis is the (network of) relation(s). To construct a network one observes the ties between actors. What can be observed as a tie in discussion groups? A formal definition of a relation is a collection of ties of a specific kind among members of a specific group (Wasserman & Faust 1994, 20). A reply from one actor to another actor is a tie, while the set of reply to and fro are considered a relation between the actors. A relation between actors is then conceptualised as an explicit communicative relation between two people based on a set of ties (i.e. the reply to one another).[1] The question is whether an initial post by someone constitutes a tie. Although it is directed to any readers of the newsgroup, it is not explicitly addressed or linked to another actor. An important advantage in identifying direct relations between contributors is that it allows for creating directed and valued graphs. An example is whether contributor A replies to contributor B, or the other way (directed), or the number of times A replies to B (valued) (cf. Wasserman & Faust 1997). In Figure 1, the organisation of a discussion as a thread is visualised.

Figure 1: A Discussion as a Thread of Contributions



Note: The letters A through E refer to actors. Squares denote messages. Subscripts refer to particular messages. The lines refer to message responses.

Based on the sequential nature of the discussion, an adjacency matrix identifying the communicative ties between actors can be constructed (Table 1). Subsequently, so-called graphs, representing the communication networks, can be constructed. Four examples of these graphs, based on the initial thread, are presented in Figure 2.

Table 1: Adjacency Matrix

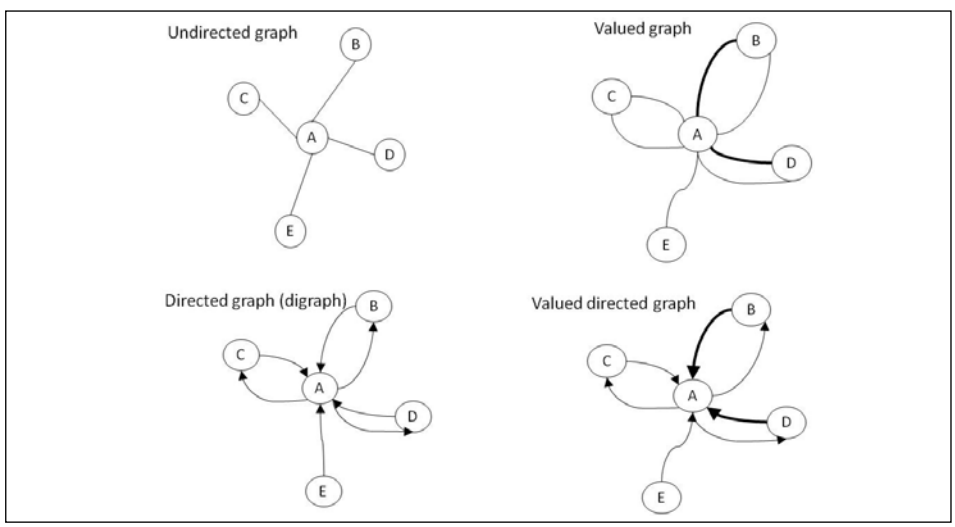
	A	B	C	D	E
A	-	0	1	1	0
B	2	-	0	0	0
C	1	0	-	0	0
D	2	0	0	-	0
E	1	0	0	0	-

Note: the letters A through E refer to actors. The numbers in the cells refer to the number of replies.

These relations between actors can be visualised in a graph. A graph is a model of a social network with (un)directed relations (cf. Iacobucci 1997). Relations can also be valued in terms of specific indicators. For instance, the number of replies directed to another contributor: the more replies, the higher the value. Also, using mixed methods, the value of a relationship might be established through content analysis of a reply. For instance, depending whether the replier agrees or disagrees with the former contributor, the relation is valued as +1 or -1.

Graph theory is important since it allows for the calculation of indices for the structure of the social network. Although there are many such indices, only a few will be mentioned. Centrality refers to the degree the network is cantered around a single actor. Density refers to the degree of interconnectedness, the level of linkage among the points in a graph (Scott 2000; Wasserman & Faust 1997). The in-degree of an actor is the number of relations that arrive at that specific actor and is an indication of its receptivity or its popularity. Or in the case of discussions, the degree the actor stimulates a discussion. It might be interpreted as opinion leadership (cf. Gao et al 2005). The out-degree refers to the number of relations depart from an actor. This is an indication of reactiveness.

Figure 2: Graphs of Communication Networks



Note: The letters A through E refer to actors. Squares denote messages; circles denote actors. The lines and arrows denote communicative relations. Lines denote undirected relations, arrows denote directed relations. The thickness of the lines denotes the value of a relation.

Although the example refers to Usenet discussions, since discussions in newsgroup reader software are often presented in a hierarchical way (see figure 1), this approach is of course applicable to all types of online discussions. Tremayne (2007b) distinguishes the network aspect of the blogosphere as a key characteristic. Characteristics he distinguishes are the textual and therefore archivable nature of blogs allowing for reconstructing the flow of ideas. The social ties, reflected in the set of links (e.g. blog roll) provided to other blogs or direct links to other blog posts.

Comparative Analysis on Discussion Groups

Discussion groups can be studied by themselves, viewing them as a specific case. Although, this may be insightful, comparing different discussion groups may prove more informative. There are several options for comparison. One option is to compare discussion groups devoted to different political parties. For instance, a discussion group devoted to political party A with a discussion group devoted to party B. Research questions could focus on whether discussions in one group are on different topics, on a more diverse range of topics and whether the groups differ in terms of cohesion or conflict. A second option is to compare discussion groups devoted to different political and societal issues. For instance, how does a discussion group on abortion compare to a discussion group on taxes? A third option is to compare discussion groups cross-nationally, taking into account (a) the different political systems as a contextual element and (b) the organisation of discussion groups. One possible angle of comparison is whether a specific type of party system (bi-party or multi-party) is reflected in the available discussion groups? For instance, in the Netherlands a large number – 24 – of political parties participated in recent national elections. However, there is only one general Usenet discussion group on politics that has no a priori political ideological stance or focus on a particular issue. In the U.S., which is primarily a two-party system, several specific groups are available for democrats and republicans. Furthermore, a large number of discussion groups contain the term 'politics'. Mostly, these newsgroups represent a clear ideological political side.

A cross-national comparative question might be whether political discussion groups where participants from all parties join in one group is different from a group where only likeminded people are (democrats or republicans amongst each other). Does this affect the communication patterns? And does this affect the content and the tone of these discussions? It is imaginable that people in a single discussion group with divergent opinions state their opinions more polarised than they would have done amongst likeminded people (cf. Tajfel & Turner 1979). This may even result in cliques of people sharing the same opinions; on the other hand, communication patterns may be predominantly between people with different opinions, not similar opinions. It is thought that disagreement in opinions is important driving force for electoral change and therefore very important in a democratic society: "If people do not encounter disagreement as part of social interaction and political communication, the deliberative efficacy of political communication is seriously compromised. Just as important, the capacity of citizens to render political judgement is fundamentally undermined" (Huckfeldt, Johnson & Sprague 2004, 8; cf. Sharon 2007).

If multiple Usenet political discussion groups are available, a further question could focus on cross-reference messages, meaning whether people decide to post

one message simultaneously to more than one group. Cross-referencing results in an augmentation of the communication network outside of a single discussion group. As such these cross-references might function as weak ties between distinct communication networks (cf. Granovetter 1973).

A final dimension for comparison has already been addressed, namely analysis of discussion groups across time. This may be before and after the elections, or comparing discussions on elections throughout the years. As such there seem to be ample opportunities to go beyond the analysis of a single case.

Issues Concerning Sampling

A cautionary remark on the availability of the data: one has to keep in mind that Usenet data are not archived on servers indefinitely. Depending on the level of activity of a discussion group and administrator policy, discussions are purged from the server periodically. More active discussion groups are archived for a smaller period than inactive groups. This retention period may differ for each internet service provider. Google groups claims to have an extensive archive in the sense that they go far back in time. However, some header data are not available. Also, since it is web based, archiving will be extremely arduous. Standard web archiving software (e.g. Teleport or Httrack) for downloading websites does not work with Google Groups. Retrieving newsgroup discussions from Usenet servers is easier since there are numerous software applications available for this task (e.g. Lurker32, Newsman Pro).

Content Analysis

Sampling from discussion groups is relatively straightforward (cf. Krippendorf 2004; Neuendorf 2002). Krippendorf distinguishes sampling units, recording/coding units and context units. An important limitation in sampling on messages in discussion groups is that messages often only can be interpreted correctly in the context of the discussion. The preferable strategy is to sample threads and use individual messages as a recording unit.

Depending on the research question, one can decide not to focus on the discussion, but on specific actors participating in the discussions. In that case the sampling unit would be the actor, and the recording unit would be the posted message.

Since the content of discussions are archived digitally, computer assisted content analysis might seem to be a logical alternative (cf. Skalski 2002). However, since discussions are the result of everyday discussions by people that do not use language in a professional manner, the usage of words is more varied and unpredictable than when the texts would have been produced by media professionals (cf. Rife, Lacy & Fico 2005). In our study, we decided to code the messages manually. Students were instructed in using a straightforward coding scheme and coded whether political actors, political parties and political themes were present in messages. To limit the amount of coding, a systematic sample (i.e. every fifth discussion) was drawn.

Social Network Analysis

Sampling is common practice in large scale data analysis. Sampling reduces the costs of data collection and allows for statistical testing to determine whether relations between variables are statistically significant. However, traditional sampling

theory does not hold in the case of network analysis, where the relation between two individuals is the unit of analysis and not the individual. If people are selected randomly, the sample of people is representative for the population. However, the relations between actors in the sample are most likely not representative for the relations between actors in the population (Scott 2000; Wasserman & Faust 1994). Only basic indices are defined at the individual level (in the case of ego networks), such as the average number of friends in individual reports, are unbiased. Therefore, for the measure of complete networks, other strategies than random sampling strategies are required.

Typically, social network analysis strives to chart complete networks. The question is what delineates a discussion network from other ones. The first boundary is the discussion group itself: only participants' relations within one or more discussion groups are of interest. If it turns out that actors cross-post messages simultaneously to other discussion groups, a researcher might choose to extend the boundaries to those groups. A second type of boundary is the time span for which the relations are being identified. Is one week of discussions enough to chart all relations between contributors or does one need a larger time span.

The second commonly used sampling strategy is the reputation method (e.g. snowballing, see Wasserman & Faust 1994). This entails an initial small sample of individuals. For these individuals is determined how they are linked to other individuals. This is repeated a number of times, until the number of additional individuals decreases steeply. This procedure seems applicable to online discussions. However, snowballing as a sampling strategy itself tends to result in biased estimation of connectedness: people with more connections will be able to name more other people than people with fewer connections. Whether this bias also occurs in discussion groups is not clear.

Dutch National Elections of 2006: Online Political Discussions in nl.politiek

In this section results are presented from analysis on discussions in the Dutch newsgroup on politics in the context of the national elections of 2006.

Sampling

All messages were downloaded from the server news.surfnet.nl from the newsgroup nl.politiek (nl.politics) in the period from 25/10/2006 to 12/25/2006 (i.e. 62 days). 1802 actors produced 2,656 discussions consisting of in total 115,793 messages. On average, this results in 1868 posts in 43 discussions per day. These data were used for content analysis, network analysis and the longitudinal analysis. From the population of all discussions, a random sample was drawn for content analysis, resulting in 214 discussions. For network analysis and the longitudinal analysis, the complete dataset, spanning eight weeks, was used.

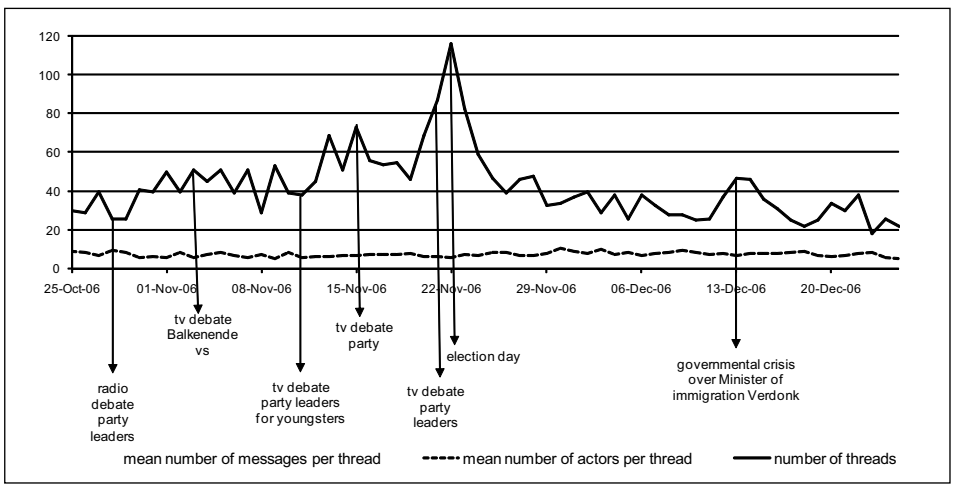
Results: Longitudinal Analysis

The main questions in the longitudinal analysis are to what extent the activities (e.g. the number of actors, the number of messages) in the discussion group varies during the election campaign and after the elections, and how political events influence political discussions on the net? In Figure 3, we observe that the

number of threads increases towards Election Day. After Election Day, the number of discussions decreases again. The number of actors is relatively constant. These results indicate that increasing activity seems to increase the number of threads, but at the same time shorten the discussion threads.

Looking at the radio and television debates in which party leaders participated, we see a weak relation with the daily activity in the newsgroup. Apparently these media events do not trigger additional discussions and participation within the newsgroup as compared to other days. On December 13 2007, we see a significant rise in activity in the discussions. On this day, the government entered a crisis when the parliament lost trust in the Minister of Immigration Verdonk. The televised parliamentary debate lasted throughout the night.

Figure 3: Trends of Participation in the Political Newsgroup nl.politiek, October 25, 2006 – December 25, 2007



These results substantiate the results reported in Table 2. Here, we see that the mean number of messages per day and the mean number of actors per discussion per day are correlated very strongly ($r=.86$, $p<.001$): the more actors participate in the discussions, the more messages they generate. This seems to be logical, although the relation is not perfect. Furthermore, we see that the more actors participate within one discussion, the less discussions threads are started ($r = -.45$, $p<.001$). This also implies that when a participant does not see reason to participate in an already existing thread, he or she most likely will start a new thread of his/her own. The number of threads per day affects the length of threads per day. The more threads are started in one day, the shorter the threads seem to become ($r = -.46$, $p<.001$).

Table 2: Correlations between Participation Indices

	mean number of messages per discussion per day	mean number of actors per discussion per day
mean number of actors per discussion per day	.861*	-
number of threads per day	-.456*	-.448*

n=62, * p < .001

Results: Content Analysis

For the results from the content analysis, see Table 3 and Table 4. The Socialist Party was the topic of discussion relatively often (33.6%), with the Christian Democrats and the Social Democrats in the second and third place respectively. These percentages for occurrence in the discussion group differ from the percentages for electoral votes. The Socialist Party, the main party in the discussion, was in third place in the elections. The other parties in the discussion group are ordered similarly to the ranking based on the electoral results. The interpretation however may be quite different. The ranking of the electoral outcome is indicative for party preference, while the ranking of political parties in the discussion group may also be interpreted in terms of party dislike. Opinions expressed in this discussion group often rather negative. As such, negative statements on specific political parties puts them higher in ranking of occurrences, but as such should be interpreted as an indication of unpopularity.

Table 3: Most Frequently Discussed Political Parties in Discussion Threads

	Occurrence in discussions		Electoral votes	
	%	Rank	%	Rank
Socialist party	33.6	1	16.6	3
Christian democrats	19.7	2	26.5	1
Social democrats	15.9	3	21.2	2
Liberal party	9.8	4	14.7	4
Green party	3.3	5.5	4.6	5
Christian party	3.3	5.5	4.0	6
Right wing parties	5.6	7	6.6	7
Democrats	0.9	8	2.0	8

n=214

Table 4, presents the most frequently discussed topics in the discussion group. The most discussed topics are "politics in general" (28%). This general category contains all posts that deal with politics without reference to a political issue, such as the subsequent topics show. In second place is "immigration policy" with 7.3 percent off all discussions. In general, the percentages are quite low, since the discussion group is solely devoted to politics. This finding corroborates earlier findings from Schneider (1997). We have two explanations for this low percentage. First of all, a narrow definition of politics – referring only to institutionalised power structures for society at large – was used. Furthermore, the absence of a moderator, someone who filters unwanted behaviour, such as insults, flaming and trolling, leads to many off-topic discussions.

Table 4: Most Frequently Discussed Topics for all Online Discussions

	Percentage of on-topic discussions
Politics in general	28.0
Immigration policy	7.3
Crime and safety	4.2
Media	3.5
Democracy	3.4
Human interest	2.8
Economy	2.5
Europe	2.0
Education	1.5
Culture	1.2
Finance	1.0
Health care	.8
Ethics, norms and values	.5
Social security	.3
Traffic and transportation	.2
Political campaign	.2
Defence and military	.0

n=214

Results: Network Analysis

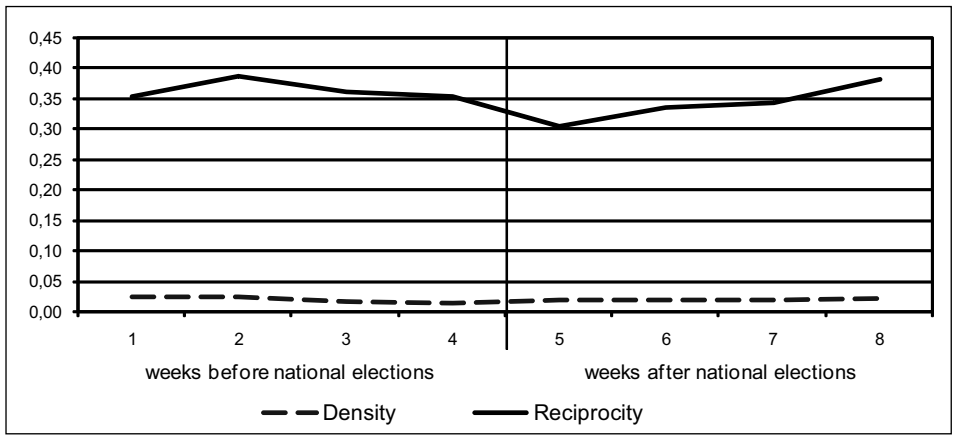
For the network analysis, two indices were calculated for each week prior and after the elections: density and reciprocity. The density of the network refers to the degree of dyadic connections in the network. In terms of binary data, density is the ratio of the number of adjacencies divided by the number of possible pairs. The degree of dyadic reciprocity is defined as the proportion of reciprocal ties of all existing ties in the network. The results aggregated for discussions per week are presented in Figure 4.

The density of the network per week is quite low, but more importantly, the density seems to be fairly constant. Also, there is no clear break around the election date (week 4 versus week 5). The degree of reciprocity also seems constant. However, as can be seen from Figure 4, the degree of reciprocity is substantially lower around the election date. This is consistent with the earlier findings that the length of the discussions decreased around Election Day. This seems to imply that during these days, contributors are more prone to state their own opinion than to engage in discussions of others.

Concluding Remarks

In this article, research strategies applicable to the study of online discussions in general and political discussions in particular were presented and demonstrated. They show that a number of very different research strategies can be applied on the raw data of online political discussions, allowing for detailed descriptions of online political discussions. Here, we only presented a few basics findings. Future research will focus on more elaborate analyses where different types of data (i.e.

Figure 4: Trends in Network Characteristics



Note: Network indices were calculated with Ucinet 6 for Windows (v6.153) (Borgatti, Everett & Freeman 2002).

network data and content analysis data) will be combined. In the remainder of this section some opportunities and difficulties on online political communication are reviewed.

Methodology

The analysis of online discussions has high potential. Data are easily accessible, in digital form and allow for unobtrusive measurement of people's opinions and actions. Using elaborate schemes for organising and analysing data, detailed information is obtained about contributor actions in discussion groups in general and here, in particular, on politics in the context of national elections. The combined use of content analysis, longitudinal analysis and network analysis allows for detailed description of the nature of online political discussions. Although there are important benefits, analysing discussion groups is not without difficulty.

Online discussion groups contain so much information and thus can be viewed from different angles – such as longitudinal analysis, network analysis, and content analysis – the combined use of research strategies is appealing. However, combining multiple strategies can prove difficult without an adequate theoretical perspective guiding the research. Furthermore, such combined research designs require complex analytical schemes and data structures; it also needs to be guided by an explicitly formulated theory. When a theory is absent and has to be developed a mixed methods approach, where qualitative and quantitative strategies are combined, may prove helpful.

Concerning network analysis, it ideally focuses on the full network. This entails the analysis of the total population of actors within pre-defined boundaries. Combining this with content analysis may prove difficult. The vast amount of content data to analyse requires computer aided content analysis. However, computer aided content analysis on non-professionally produced content requires insight in the informal vocabulary of everyday speech. Although it is not impossible, it is far more elaborate than with professionally produced content.

Another important issue is the occurrence of missing observations. One type of observation that cannot be accounted for is participation in discussion groups by lurkers. Lurkers are people that only read, but not contribute to the discussing group and are therefore not observable. How large this group is, is not clear. Only in discussion group where one has to subscribe, it is possible to gain some insight in the lurkers issue. The non-contributors may lurk the group, or may not read group discussions. Without obtrusive measurement this remains unclear. As such, participation in online discussions equals contributing to these discussions. Many software packages for network analysis and computer assisted content analysis have been developed; for an overview of social network software, consult Huismans and Van Duin (2005). There are some attempts to design specialised software applications for communication network analysis on the net, for instance tapping into data of social network sites such as Vizster (Heer & Boyd 2005), predominantly focusing on visualising online social networks (e.g. Friendster), they lack analytical and numerical features of generic SNA software such as Ucinet (Borgatti, Everett & Freeman 2002) and Pajek (De Nooy, Mrvar & Batagelj 2005).

Extracting relational data from messages can be done by hand, through coding. However, given the large data quantities in discussion groups, a more automated procedure is advisable. Since downloaded discussions consist primarily of unordered textual data, "regular expressions"[2] were used to delete redundant information, leaving us with network data. For an overview on automated content analysis, consult Skalski (2002).

Theory

Whether online discussions lead to more political participation and empowerment of peripheral groups, requires further empirical investigation. It is clear that online discussions will gain more possibilities and opportunities for participation in politics, governance and society. How it will develop in countries with different political and media systems is a question still unanswered.

More generally, viewing participating in virtual communities as a form of converging or diverging opinions (cf. Rogers & Kincaid 1981) is a manifestation of underlying conflicts of interests (cf. Coser 1956). Applied to online political discussion, people vary with respect to their party preferences. These party preferences are related to different interests people have, identifiable by explicitly mentioned party preferences, by mentioning political parties in favourable words, or by certain issues linked to specific political parties. It is not clear what role these conflicting interests play in the development of specific discussions and what it means for the discussion group as a whole. For example, is the degree of conflict related to the degree of density within a network, implying that conflict stimulates the activity in discussions? Also, does it differ for general discussion groups defined in terms of a general issue (i.e. heterogeneous group composition), as compared to a discussion group formed around an issue that is defined very narrow (i.e. homogeneous group composition).

Finally, content analysis of discussion groups offers possibilities to see whether the public agenda (cf. McCombs & Shaw 1972) corresponds to the media and/or the political agenda. Because the discussions in newsgroup reflect daily opinions in the newsgroup, detailed causal analysis can be performed as to whether there

is a correlation between the agenda and what agenda precedes the other (cf. Kleinnijenhuis et al 2007).

Notes:

1. Relations are often also referred to as lines, edges or arcs. Actors are also referred to as nodes, vertices or points.

2. Regular expressions (Friedl 2006) allows for complex search procedures, using an array of wildcards.

References:

Borgatti, Stephen, P., Martin G. Everett and Linton C. Freeman. 2002. *Ucinet for Windows: Software for Social Network Analysis*. Harvard, MA: Analytic Technologies.

Boyd, Danah M. and Ellison, Nicole B. 2007. Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*, 13, 1, 210-230.

Brants, Kees. 2005. Guest Editor's Introduction: The Internet and the Public Sphere. *Political Communication* 22, 2, 143-146.

Brants, Kees and Philip van Praag. 2006. Signs of Media Logic Half a Century of Political Communication in the Netherlands. *Javnost – The Public* 13, 1, 25-40.

Burnett, Gary and Laurie Bonnici. 2003. Beyond the FAQ: Explicit and Implicit Norms in Usenet Newsgroups. *Library & Information Science Research* 25, 3, 333-351.

Coser, Lewis. 1956. *The Functions of Social Conflicts*. New York: Free Press.

Creswell, John W. and Vicky L. Plano Clark. 2007. *Designing and Conducting Mixed Methods Research.* London: Sage.

Dahlgren, Peter. 2002. In Search of the Talkative Public: Media Deliberative Democracy and Civic Culture. *Javnost – The Public* 9, 3, 5-26.

Dahlgren, Peter. 2005. The Internet, Public Spheres, and Political Communication: Dispersion and Deliberation. *Political Communication* 22, 2, 147-162.

Davis, Richard. 2005. *Politics Online. Blogs, Chatrooms and Discussion Groups in American Democracy.* New York: Routledge.

De Nooy, Wouter, Andrej Mrvar, and Vladimir Batagelj. 2005. *Exploratory Social Network Analysis with Pajek*. Cambridge: University Press.

Drew, Paul and John Heritage. 2006. *Conversation Analysis*. London: Sage.

Friedl, Jeffrey. 2006. *Mastering Regular Expressions*. Farnham, Surrey: O'Reilly.

Galtung, Johan. 1967. *Theory and Methods of Social Research*. London: Allen & Unwin.

Gao, Jumbo, Min Zhang, Fan Jang and Xuja Wang. 2005. Analysis of Opinion Leader in Online Communities. *Knowledge-based Intelligent Information and Engineering Systems*. Berlin: Springer. <http://www.springerlink.com/content/bxrhdnvtrhd2e2vr/?p=915da1a24463400595451a9c1904a68b&pi=0>

Goffman, Erving. 1959. *The Presentation of Self in Everyday Life*. London: Pelican Books.

Google. 2001. *Google Acquires Usenet Discussion Service and Significant Assets from Deja.com*. Google Press Release. <http://www.google.com/press/pressrel/pressrelease48.html>

Hallin, Daniel C. and Paolo Mancini. 2004. *Comparing Media Systems. Three Models of Media and Politics*. Cambridge: Cambridge University Press.

Hill, Kevin A. and John E. Hughes. 1997. Computer-Mediated Political Communication: The USENET and Political Communication. *Political Communication* 14, 1, 3-27.

Huckfeldt, Robert, Johnson, Paul E. and Sprague, John. (2004). *Political Disagreement. The Survival of Diverse Opinions within Communication Networks*. Cambridge: Cambridge University Press.

Huisman, Mark, and Marijtje A. J. Van Duin. 2005. Software for Social Network Analysis. In P.J. Carrington, J. Scott, and S. Wasserman (eds.), *Models and Methods in Social Network Analysis,* 270-316. New York: Cambridge University Press.

Iacobucci, Dawn. 1997. Graphs and Matrices. In S. Wasserman and K. Faust (eds.), *Social Network Analysis. Methods and Applications,* 92-166. Cambridge: Cambridge University Press.

Jankowski, Nicholas, W. 2002. Creating Community with Media: History, Theories and Scientific Investigations. In L. Lievrouw and S. Livingstone (eds.), *Handbook of New Media: Social Shaping and Consequences of ICTs,* 34-49. London: Sage.

Kleinnijenhuis Jan, Anita M.J. van Hoof, Dirk Oegema, and Jan de Ridder. 2007. A Test of Rivaling Approaches to Explain News Effects: News on Issue Positions of Parties, Real-World Developments, Support and Criticism, and Success and Failure. *Journal of Communication* 57, 2, 366-384.

Krippendorf, Klaus. 2004. *Content Analysis: An Introduction to its Methodology 2nd* ed. Thousand Oaks, CA: Sage.

McCombs, Maxwell. E. and David L. Shaw. 1972 The Agenda-setting Function of Mass Media. *Public Opinion Quarterly* 36, 2, 176–187.

Meraz, Sharon. 2007. Analyzing Political Conversation on the Howard Dean Candidate Blog. In M. Tremayne (ed.), *Blogging, Citizenship and the Future of Media* 59-81. London: Routledge.

Monge, Peter R. and Contractor, Noshir S. 2003. *Theories of Communication Networks*. Oxford: University Press.

Muhlberger, Peter. 2004. Access, Skill, and Motivation in Online Political Discussion: Testing Cyberrealism. In P.M. Shane (ed.), *Democracy Online. The Prospects for Political Renewal through the Internet,* 225-238. New York: Routledge.

Neuendorf, Kimberly A. 2002. *The Content Analysis Guidebook*. New York: Sage.

Norris, Pippa. 2001. *Digital Divide: Civic Engagement, Information Poverty, and the Internet Worldwide.* NewYork: Cambridge University Press.

O'Sullivan, Patrick B. and Andrew F. Flanagin. 2003. Reconceptualizing 'Flaming' and other Problematic Messages. *New Media & Society* 5, 1, 69-94.

Papacharissi, Zizi. 2004. Democracy Online: Civility, Politeness, and the Democratic Potential of Online Political Discussion Groups. *New Media & Society* 6, 2, 259-283.

Park, Han Woo. 2003. Hyperlink Network Analysis: A New Method for the Study of Social Structure on the Web. *Connections* 25, 1, 49-61.

Rife, Daniel, Stephen Lacy and Frederick G. Fico. 2005. *Analysing Media Messages*. London: Lawrence Erlbaum Association.

Rogers, Everett, M. 1995. *Diffusion of Innovations* 4th ed. New York: Free Press.

Rogers, Everett, M. and D. Lawrence Kincaid. 1981. *Communication Networks. Towards a New Paradigm for Research*. New York: Free Press.

Sack, Warren. 2001. Conversation Map: An Interface for Very Large-Scale Conversations. *Journal of Management Information Systems* 17, 3, 73-92.

Savigny, Heather. 2002. Public Opinion, Political Communication and the Internet. *Politics,* 22, 1, 1-8.

Schneider, Steven M. 1997. *Expanding the Public Sphere through Computer-Mediated Communication: Political Discussion about Abortion in a Usenet Newsgroup*. Unpublished PhD thesis. Massachusetts: MIT

Scott, John. 2000. *Social Network Analysis* (2nd ed.). New York: Sage.

Shane, Peter M., ed. 2004. *Democracy Online. The Prospects for Political Renewal Through the Internet*. New York: Routledge.

Skalski, Paul D. 2002. Computer Content Analysis Software. K.A. Neuendorf (ed.), *The Content Analysis Guidebook,* 225-239. New York: Sage.

Smith, Marc.1999. Invisible Crowds in Cyberspace: Measuring and Mapping the Social Structure of USENET. In M. Smith and P. Kollock (eds.), *Communities in Cyberspace: Perspectives on New Forms of Social Organization* 196-219. London: Routledge.

Tajfel, Henri and Turner, John, C. 1979. An Integrative Theory of Intergroup Conflict. In W.G. Austin and S. Worchel (eds.), *The Social Psychology of Intergroup Relations*, 33-47. Monterey, California: Brooks/Cole.

Tedesco, John C. 2004. Changing the Channel: Use of the Internet for Communicating about Politics. In L.L. Kaïd (ed.), *Handbook of Political Communication Research,* 507-532. Mahwah: Lawrence Erlbaum.

Tremayne, Mark, ed. 2007a. *Blogging, Citizenship and the Future of Media*. London: Routledge.

Tremayne, Mark. 2007b. Introduction, The Blog-Media Relation. In M. Tremayne (ed.), *Blogging, Citizenship and the Future of Media*, ix-xix. London: Routledge.

Tremayne, Mark, Nan Zheng, Jae Kook Lee and Jaeong Jeong. 2006. Issue Publics on the Web: Applying Network Theory to the War Blogosphere. *Journal of Computer-Mediated Communication 12,* 1, 290-310.

Turner, Tamara Combs, Marc A. Smith, Danyel Fisher and Howard T. Welser. 2005. Picturing Usenet: Mapping Computer-Mediated Collective Action. *Journal of Computer-Mediated Communication* 10, 4, article 7. <http://jcmc.indiana.edu/vol10/issue4/turner.html>

Van Os, Renée, Nicholas W. Jankowski, and Maurice Vergeer. 2007. Political Communication on the Internet during the 2004 European Parliament Election Campaign in Nine EU Member States. *European Societies* 9, 4, 565-585.

Wasserman, Stanley and Katherine Faust. 1997. *Social Network Analysis. Methods and Applications*. Cambridge: Cambridge University Press.

Wellman, Barry. 2001. Computer Networks as Social Networks. *Science* 29, 3, 2031-2034.