

Discussion catalysts in online political discussions: Content importers and conversation starters

Itai Himelboim

101L Journalism Building, University of Georgia, Athens, GA, 30602

Eric Gleave

Marc Smith

17950 Preston Rd., Suite 310, Dallas, Texas 75252

This study addresses 3 research questions in the context of online political discussions: What is the distribution of successful topic starting practices, what characterizes the content of large thread-starting messages, and what is the source of that content? A 6-month analysis of almost 40,000 authors in 20 political Usenet newsgroups identified authors who received a disproportionate number of replies. We labeled these authors “discussion catalysts.” Content analysis revealed that 95 percent of discussion catalysts’ messages contained content imported from elsewhere on the web, about 2/3 from traditional news organizations. We conclude that the flow of information from the content creators to the readers and writers continues to be mediated by a few individuals who act as filters and amplifiers.

doi:10.1111/j.1083-6101.2009.01470.x

Political discussions are key elements of democratic societies in which citizens are expected to make informed decisions with regard to issues of civic importance (Baker, 1989). Traditional mass media like TV, radio, and print have all played an important role in distributing information and informing opinions (Picard, 1985). Researchers have recognized that a few influential individuals play a critical role in mediating this flow of information between mass media and the public (Lazarsfeld, Berelson & Gaudet, 1948; Katz & Lazarsfeld, 1955). Computer-mediated discussion tools may shift the nature of political discussion by including a wider variety of perspectives and voices, but do they change patterns of information flow?

This study explores the flow of information among participants in online political discussion forums and the sources of information found in online forum content. Specifically, we are interested in those who start discussions that attract many participants and messages. These contributors potentially play a unique role in

shaping the discussion group's topic agenda. We ask a set of core questions: What is the distribution of successful topic starting practices in online political discussions? What characterizes the content of messages that start large threads? And, What is the source of that content?

A social roles perspective identifies participants who attract more replies than others and thus potentially disproportionately influence discussion. We report on the analysis of nearly half a million messages in 20 political newsgroups in 6 month-long periods. A small number of authors received the most replies among all those who started new discussion threads. Criteria based on distributions of thread characteristics identify the population of high reply-attracting participants, a group of contributors we label "discussion catalysts." We then analyze the content of their messages to explore common elements of text that attracted large numbers of replies. Content analysis added additional insight into the nature of the material and information sources that attracted many replies.

Political discussions, Internet and Society

Forums for political discussion play a critical role in societies and, in particular, democracies (de Toqueville, 1945 [1839]). Political discussions have shifted in large numbers to computer-mediated discussion spaces like Usenet newsgroups, web boards, and e-mail lists (Levine, 2000). These online conversations may be banal but focus on issues of civil importance. Informed citizens are crucial for the effective operation of a democratic society (Siebert, Peterson, and Schramm, 1963; Picard, 1985). The Internet has been a cause for great enthusiasm among those who appreciate that it gives a large population access to a wide range of sources of information and gives each person the potential to create new information or nominate new topics for public discussion. Corrado and Firestone (1996: 17) believed that online discussions, especially Usenet, will create a conversational democracy where "citizens and political leaders interact in new and exciting ways." Hauben and Hauben (1997) suggested that online discussion groups allow citizens to participate within their daily schedules. Rheingold declared that if discussion boards are not democratizing technology, "there is no such thing" (1993: 131). Business leaders invoke the Internet as a guarantee of press freedom (McMillan, 2008). Researchers have documented the potential for effective collective action through the Internet (Bucy & Gregson, 2001; Mehra, Merkel, & Peterson 2004; Kahn & Kellner 2004). Because computer-mediated discussions provide an almost infinite canvas for new messages, the scarce resource becomes attracting attention, particularly replies to new threads. Reply counts are therefore a useful indicator of the value of or interest in that topic.

In the following, we first draw from the mass communication literature to examine the ways in which information flows into and through computer-mediated political discussions. We then draw from the sociology of roles to identify the

critical social practices that are core to the activity of computer-mediated discussion groups.

The 'two-step flow' theory proposed by Lazarsfeld, Berelson, and Gaudet (1948) emphasized the role of the individual in the vertical flow of information from mass media sources to particular members of the potential audience. Katz and Lazarsfeld (1955) recognized the importance of a few highly connected opinion leaders in the flow of information from mass media to individuals. The Rovere study (Merton, 1957) was one of the first studies designed to investigate the personal influence opinion leaders had on individuals on a range of topics such as politics, fashion, and movies. A few individuals were found to be more tuned to certain mass media messages related to their particular topics of interest or expertise. These individuals were also trusted more by their peers and were sought out for guidance and information. In the following decades, scholars have extended the original notion of a two-step flow by pointing to the possibility of a multistep flow, but the basic idea remains a prominent account of media effects (see Brosius & Weimann, 1996; Katz, 1987). Recently, Southwell and Yzer (2007) suggested a new implementation of the model, examining the role of the two-step flow and interpersonal communication in the context of mass media health campaigns.

Major criticism of the two-step flow theory challenged Lazarsfeld and Katz for giving legitimacy to the elites who set agendas (Gitlin, 1978; Hochheimer, 1982). Further criticism of the two-step flow model and its conception of the role of opinion leaders focus on how these models ignore the existence of many steps and their lack of attention to the directions of the flows of information. Weimann, (1994) suggests the need for a distinction between the flow of information and the flow of influence, which further demands major amendments to these models. Research to support the two-step model was based on the use of recall survey and interview data that asked people to recount all of their interactions (Van Den Ban, 1964). The resulting data was of questionable completeness and accuracy. Critics raised these data limitations as additional challenges to the two-step flow model, focusing on the impossibility, at that time, of capturing a significant portion of interaction events (Weimann, 1994). Information can leak and move through many channels making it more difficult to establish the steps the flow of information takes through a population. Without such data, it has been difficult to document the role of key leaders and their effect on information flows.

New data to support the study of these phenomena and overcome the data limitations of the model are now available from the data generated by the interactions of large populations through computer-mediated discussion systems. The dynamics created by messages and the replies they attract might not address influence or opinion change, but is a manifestation of core aspects of the flow of information.

Usenet discussion newsgroups

With the commercialization of the Internet, Usenet stands out as a remaining 'open' region for conversation and discussion. The Internet was quickly colonized by commercial interests, mainly for its potential for advertisers (Bagdikian, 2004). Some argue that capitalist patterns of production transform the Internet into a commercially oriented media that has little to do with promoting social welfare or democratic practices (Papacharissi, 2002). Usenet is an interesting contrast developed in the early 1980s as a mechanism for exchanging threaded discussion content in a 'peer-to-peer' model. Unlike a website with a web board or discussion page, no single individual or organization controls all of Usenet (Wikipedia, 2007). Usenet newsgroups are far more anarchic than web-hosted discussions in that they lack central authorities that can police or moderate. Yahoo, for example, was accused of deleting users' comments on its photo-sharing website Flickr (BBC, 2007). In contrast, most Usenet newsgroups have no moderators and thus theoretically support a free flow of information and an open market of opinions and ideas. The distributed nature of Usenet means that Usenet messages cannot be easily censored without high levels of cooperation and coordination across many jurisdictions, institutions, and individuals.

Newsgroups are a hybrid of broadcasting and interpersonal communication

Internet discussion groups create a form of conversation in which participants, by posting messages, broadcast their opinions to an entire population of potential observers and directly respond to a particular participant at the same time. A good metaphor for Usenet is an informal town meeting. These meetings take the form of a discussion when individuals respond to opinions and ideas presented by their peers. At the same time, anything said is available for everybody to hear. Within newsgroups, communications are interpersonal and simultaneously broadcast.

In Usenet, the 'town' expands to potentially include people from all over the world. Anyone with Internet access and the appropriate language and technical skills can post a message, respond to others, and aim to change the course of a discussion or evoke discussion by starting a new thread. With the exception of the thread starting messages, all posts are both a response to a specific message (and the message's author) and a broadcast to the newsgroup. Differences in the patterns of initiation and reply around each author are indications of the social roles each plays. Sociological theories of social roles are useful guides to the study of these phenomena.

Social roles in online discussions

A social role perspective highlights patterns of recurring social behavior among members of a group. These patterns indicate which social roles each participant performs. In the context of online fora, social roles are identified via patterns generated by the exchange of messages in reply to one another. Other approaches, such as content or discourse analysis, are valuable alternative approaches; however,

message-level analysis does not address the newsgroup dynamics in which some messages gain attention in part by attracting many reply messages.

Participants in online discussion groups can perform a range of social roles. Online social roles have been identified through ethnographic studies of the content of interaction (Golder & Donath, 2004; Donath, 1996; Marcocci, 2004). Some participants are experts who answer other people's questions (Welser et al., 2007). Others act as conversationalists, contributing to large threads with many messages and branches. More aggressive and confrontational roles include flame warriors and trolls who engage in or incite angry exchanges (Burkhalter & Smith 2004; Golder & Donath 2004; Turner et al., 2005; Herring, 2004; Haythornthwaite & Hager, 2005). Some effort has been made to use behavioral and structural cues to identify these roles via visualizations of patterns of initiation, reply, and thread contribution rates over time. These studies resulted in the identification of distinct patterns of contribution, which are proposed as indicators of distinct social roles (Viegas & Smith, 2004; Turner et al., 2005). These methods helped identify new roles including the role of 'question person' and 'discussion person' (for further discussion on online social roles see Welser et al., 2007).

Structural analyses can be a powerful tool for identifying social roles. Turner and Fisher et al. (2006) used local network attributes to identify four social roles in computer-mediated discussions: members, mentors, managers, and moguls. Welser et al. (2007) used structural signatures to identify answer people in technical support newsgroups.

Social roles in online political discussions

Unlike historical political discussion venues, anyone in newsgroup discussions can introduce a new issue, topic, or opinion by starting a thread or replying to another message. Those who are most able to evoke contributions to the discussion from others play a unique social role as the introducers of discussion topics. We label this social role "discussion catalysts" because of their ability to stimulate the conversation of others.

Some messages start new threads and therefore indicate the introduction of a topic for discussion. By looking closely at the patterns of reply to thread starting messages, we can focus on the particular social process of topic nomination. Authors who frequently nominate new topics perform a social role of interest to theories of political communication because of their function as filter, selector, and amplifier of particular topics.

Some individuals start new threads that disproportionately attract replies from many different repliers. Only a small number of messages and authors receive a significant fraction of all the replies. Some people do not start many or any threads but do contribute many messages, potentially to many threads. These patterns suggest that behavior in these social spaces is specialized and is indicative of possible social roles.

Participants in internet discussion spaces share information and exchange opinions side by side with a wide variety of other information platforms, such as blogs, wikis and websites of traditional news organizations. Mass communication literature suggests that a few individuals who are highly connected within their social networks mediate information flow from mass media to individuals. Are these patterns replicated on the Internet? Sociological literature guides our novel use of structural analysis to identify individuals who play unique social roles in online political discussions. We identify who plays key social roles in political discussions and provide methods for defining these social roles.

Method

Data

Our study analyzes patterns of thread initiation and reply from more than 16,000 authors in 6 months from 20 political newsgroups collected from the Microsoft Research Netscan dataset. Message content was retrieved from Google Groups. Newsgroups were selected to generate a stratified sample of all newsgroups active during July of 2006 that had 'politic' in their name. A population of 639 newsgroups was identified. To limit our focus to active newsgroups we excluded newsgroups that had fewer than 50 participants, leaving 115 newsgroups. We excluded twenty-three because they were not in English and we lacked non-English language content analysis resources. The remaining newsgroups were placed in one of four subcategories: 24 state level politics, 31 national level politics, 21 issue-specific politics and 16 non-American politics. Five newsgroups were randomly selected from each of these categories, resulting in the set in Table 1.

Within each newsgroup, we found a population of authors, each with a collection of attributes that described their activity and social structure within the newsgroup, based on their message posting behavior. Three measures about each author identified participants who started successful threads: reply share, replier share, and success ratio. A reply is defined during the message creation process by including the identifier of the message targeted for a reply in the header of the reply message. This paper uses these links between messages to define a "reply." Participants who start new threads and are high in these measures have more influence on the topics discussed with each newsgroup, as more individuals discuss the topics they bring to the table. In a sense, these participants catalyze longer discussions.

We define the reply share as the ratio of replies in threads initiated by an author to the total number of replies in the newsgroup. The replier share is the ratio of newsgroup authors who post messages in an author's threads to the total number of newsgroup participants. Finally, the success ratio is the proportion of threads an author initiates which receive replies from at least two other authors.

To identify discussion catalysts, we looked for authors who measured highly on the success ratio, reply share, and replier share measures. Because of the skewed

Table 1 Selected newsgroups, population size, and message volume

	July		August		September		October		November		December	
	Vol.	Size	Vol.	Size	Vol.	Size	Vol.	Size	Vol.	Size	Vol.	Size
ab.politics	1353	226	1073	221	1353	226	999	206	1221	218	1107	230
alt.politics.clinton	651	241	787	184	651	241	632	186	626	226	920	184
alt.politics.england.misc	593	158	610	139	593	158	391	135	304	110	510	163
alt.politics.gw-bush	5110	751	5364	766	5110	751	9716	932	8848	942	5926	740
alt.politics.homosexuality	10064	868	8894	798	10064	868	13370	1571	10166	855	7582	697
alt.politics.immigration	6878	700	4918	684	6878	700	7582	677	6408	716	7091	760
alt.politics.socialism	263	97	279	100	263	97	191	67	515	114	266	57
alt.politics.usa	10988	1423	10018	1328	10988	1423	11022	1415	9457	1293	8458	1194
alt.politics.usa.republicans	167	83	155	79	167	83	157	74	359	102	435	142
can.politics	19279	1653	16685	1604	19279	1653	15962	1637	14830	1529	16655	1517
fl.politics	1323	177	1430	146	1323	177	1864	245	2241	268	740	134
hawaii.politics	2255	120	695	113	2255	120	912	76	723	81	642	78
nyc.politics	2395	362	2422	384	2395	362	2151	391	2043	377	1563	290
nz.politics	3903	510	3910	564	3903	510	4373	505	3512	454	2542	365
scot.politics	435	48	819	103	435	48	649	54	257	43	330	37
seattle.politics	8417	503	7811	589	8417	503	10252	624	7805	563	6093	488
talk.politics.animals	447	98	968	104	447	98	278	48	596	94	667	106
talk.politics.drugs	1572	162	884	175	1572	162	431	123	1052	239	412	105
talk.politics.european-union	349	119	207	78	349	119	1751	173	1700	147	2911	139
talk.politics.medicine	677	101	1255	123	677	101	1107	145	824	129	564	90

distribution of replies and repliers, one can predict that only a few participants are defined as discussion catalysts. Operationalizing the definition of discussion catalyst requires that we set thresholds for each of these ratios.

The distribution of the success ratio is, aside from a large proportion of authors at the value one, normally distributed around 0.5. Since the values are independent of the number of an author's root messages, there are more instances of simpler fractions like 1, 0.5, and 0.25 which can be arrived at in numerous ways. To permit the capture of participants who start a very large number of threads that usually are successful, we set this threshold low at 0.5.

Reply share is highly skewed with a few individuals receiving a large value and most obtaining values near zero. We chose a threshold of two standard deviations above the mean of the logged distribution, thereby limiting ourselves to, at most, one eighth of the population (by Chebyshev's inequality). This corresponds to an unlogged value threshold of 0.14.

The threshold for replier share was constructed in the same fashion as reply share. Taking the cutoff of two standard deviations above the mean of the logged distribution produces a threshold of 0.22, again theoretically limited to, at most, one eighth of the population and, in practice, a far smaller pool.

Content analysis

Discussion catalysts post content that evokes many replies. To explore the nature of this content, 10 percent of all root messages posted by discussion catalysts were randomly selected. Each root message was coded for its source of information: traditional news organizations' websites, news websites with no presence off the web, personal websites or blogs, and websites associated with governments and NGOs. Each message was also coded for the author's original contribution: no contribution, brief contribution (up to two sentences) and substantial contribution (more than two sentences). Sources of imported information were clearly identified in messages content, either by a URL or by explicit reference to the name of the source. The content analysis was performed by one of the authors of this paper. A 10 percent random sample to assess the reliability produced overall agreement of 97% and Scott's π of 0.90.

Results

Over the 6 months from July to December, 2006, 16,513 authors created 444,643 messages in the 20 Usenet political newsgroups selected for analysis in this study. Thirteen per cent of messages initiated a thread while the remaining were replies. Only 32 percent of authors started threads. Newsgroup size ranged from 37 to 1,653 authors and newsgroup volume varied between 155 and 19,279 messages. Table 1 details the rates of population size and message volume across these newsgroups for the 6 months studied.

Table 2 Discussion catalysts per newsgroup

Newsgroup	ID	Month	Reply Ratio	Replier Ratio	Success Ratio
ab.politics	1	November	0.340	0.372	0.889
ab.politics	1	December	0.368	0.454	0.874
ab.politics	21	July	0.464	0.443	0.698
ab.politics	21	August	0.187	0.317	0.702
ab.politics	21	September	0.554	0.489	0.795
ab.politics	21	October	0.302	0.431	0.694
ab.politics	21	November	0.379	0.442	0.711
ab.politics	21	December	0.540	0.496	0.752
alt.politics.clinton	14	August	0.393	0.386	0.688
alt.politics.clinton	14	October	0.753	0.434	0.735
alt.politics.clinton	14	November	0.391	0.345	0.762
alt.politics.clinton	14	December	0.320	0.323	0.692
alt.politics.england.misc	6	July	0.713	0.683	1.000
alt.politics.england.misc	6	August	0.573	0.425	0.900
alt.politics.england.misc	6	October	0.312	0.356	0.667
alt.politics.england.misc	12	November	0.253	0.282	1.000
alt.politics.gw-bush	8	July	0.314	0.295	0.535
alt.politics.gw-bush	8	August	0.208	0.269	0.584
alt.politics.gw-bush	8	September	0.274	0.342	0.680
alt.politics.gw-bush	8	October	0.258	0.296	0.598
alt.politics.gw-bush	8	November	0.216	0.276	0.593
alt.politics.gw-bush	8	December	0.329	0.325	0.608
alt.politics.homosexuality	15	October	0.150	0.407	0.800
alt.politics.homosexuality	18	September	0.182	0.298	1.000
alt.politics.homosexuality	18	November	0.368	0.345	1.000
alt.politics.homosexuality	18	December	0.400	0.248	1.000
alt.politics.homosexuality	19	August	0.303	0.301	1.000
alt.politics.socialism	26	November	0.371	0.307	1.000
alt.politics.usa.republicans	7	December	0.682	0.824	0.769
alt.politics.usa.republicans	9	August	0.265	0.278	1.000
fl.politics	2	July	0.766	0.548	0.583
fl.politics	2	August	0.670	0.512	0.703
fl.politics	2	September	0.499	0.367	0.664
fl.politics	18	November	0.291	0.429	1.000
hawaii.politics	10	September	0.145	0.225	0.800
hawaii.politics	16	October	0.416	0.342	1.000
hawaii.politics	20	August	0.551	0.442	1.000
hawaii.politics	25	November	0.436	0.333	0.609
hawaii.politics	27	December	0.591	0.432	0.933
scot.politics	3	November	0.160	0.256	1.000
scot.politics	4	September	0.267	0.292	0.800
scot.politics	6	July	0.358	0.301	0.833
scot.politics	6	August	0.525	0.302	0.857
scot.politics	6	December	0.393	0.368	1.000
scot.politics	17	August	0.148	0.223	0.667
seattle.politics	18	November	0.267	0.224	0.875
talk.politics.animals	23	July	0.248	0.233	1.000

Table 2 (Continued)

Newsgroup	ID	Month	Reply Ratio	Replier Ratio	Success Ratio
talk.politics.drugs	5	July	0.171	0.226	1.000
talk.politics.drugs	5	September	0.367	0.332	1.000
talk.politics.drugs	5	November	0.332	0.297	1.000
talk.politics.drugs	5	December	0.637	0.253	0.600
talk.politics.drugs	22	October	0.469	0.317	1.000
talk.politics.european-union	11	October	0.268	0.295	0.557
talk.politics.european-union	13	November	0.546	0.449	0.508
talk.politics.european-union	24	October	0.286	0.243	0.595
talk.politics.medicine	28	September	0.321	0.327	0.579

Only a few thread starting authors were able to attract a significant number of replies and repliers to their messages. A small number of authors were at the far end of the distributions for the three ratios that captured their ability to initiate threads that attracted responses. Among this population of contributors, only 28 authors (0.17 per cent of all authors; 0.53 percent of all thread starting authors) were identified above the thresholds for the success, replier share, and reply share ratios. These 28 authors created 7,032 new threads (12 percent of all threads) which then attracted 88,129 messages (23 percent of all messages). These threads attracted a similarly large fraction of repliers; 4,904 authors (30 percent of all authors) responded to their threads. See Figure 1. Of this population, most authors replied to both discussion catalysts and other thread starting authors, which is why the combined author count exceeds the population. Authors identified as discussion catalysts appeared in 15 of the 20 newsgroups in this study. In five of the newsgroups—alt.politics.immigration, alt.politics.usa, can.politics, nyc.politics and nz.politics—we found no authors who met the definition of a discussion catalyst.

The distributions of reply share and replier share were skewed. Discussion catalysts' reply share ratios varied between 0.14 and 0.76. Furthermore, nine discussion catalysts were able to attract more than half of the entire population of reply messages to threads they started in a single month. The replier share ratio for discussion catalysts varied between 0.22 and 0.82. The ratio of discussion catalyst's successful threads varied from 0.51 to 1.0.

Some of the authors playing the role of discussion catalyst performed the role regularly and consistently over time. Others remained in these social spaces but dropped below the thresholds. Eight authors persisted in that role for more than a single month. One appeared in two different months; two in three months; two in four months; one in five months; and two in six months. Only two discussion catalysts were detected in more than one newsgroup.

The rate of new thread creation would seem like a good predictor for attracting other authors and their reply messages, since each additional initiated thread increases the possibility of receiving a reply. Indeed, we found that the number of threads started was correlated with the number of replies received by that author (Pearson r

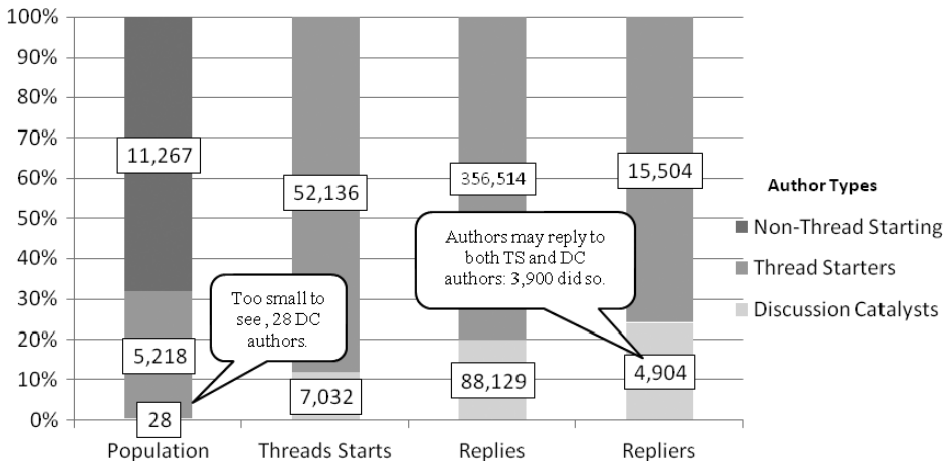


Figure 1 Relative rates of thread starting, reply attraction, and respondent attraction for three author types.

ranging between 0.63 and 0.98, $p < 0.01$). Furthermore, the number of root messages was also correlated with the number of replying authors (Pearson's R ranging between 0.48 and 0.96, $p < 0.01$). Importantly, however, the number of threads an author initiates is **not** correlated with their success ratio. This ratio separates those who collect many replies and repliers through a strategy of volume posting from those who both generate many threads and have a high rate of attracting replies and repliers. Some authors start only one successful thread, resulting in a success ratio of 1, but those authors rarely have high replier and reply ratios. In this way, discussion catalysts are distinguished by the fact that they both generate more threads and attract significant volumes of response.

The patterns of connections between two discussion catalysts and their replier populations can be seen in figures 2 and 3. These network graphs represent each author as a dot. Each reply from one author to another is represented as a line with an arrow in the direction of reply. The size of each dot represents the number of inward connections (in-degrees) each author has. The thickness of each line is proportional to the number of replies sent from that author to the other. The central role of discussion catalysts is visible: they sit within a network of connections from other authors as seen by the number of arrowheads pointing inwards towards them. This pattern is in contrast to other roles, like the answer person (Welser et al., 2007), who feature large number of outward arrows and few inward pointers. The figures illustrate the distinction of discussion catalyst behavior from other participants.

Figure 2 illustrates the social structure of a political discussion newsgroup, ab.politics, and highlights the presence of two key contributors, located in the two opposite corners, who attracted far more replied messages than any other contributors to the newsgroup. Figure 3 shows a similar pattern in the alt.politics.england.misc newsgroup. Here, however, only a single participant, located in the center, has

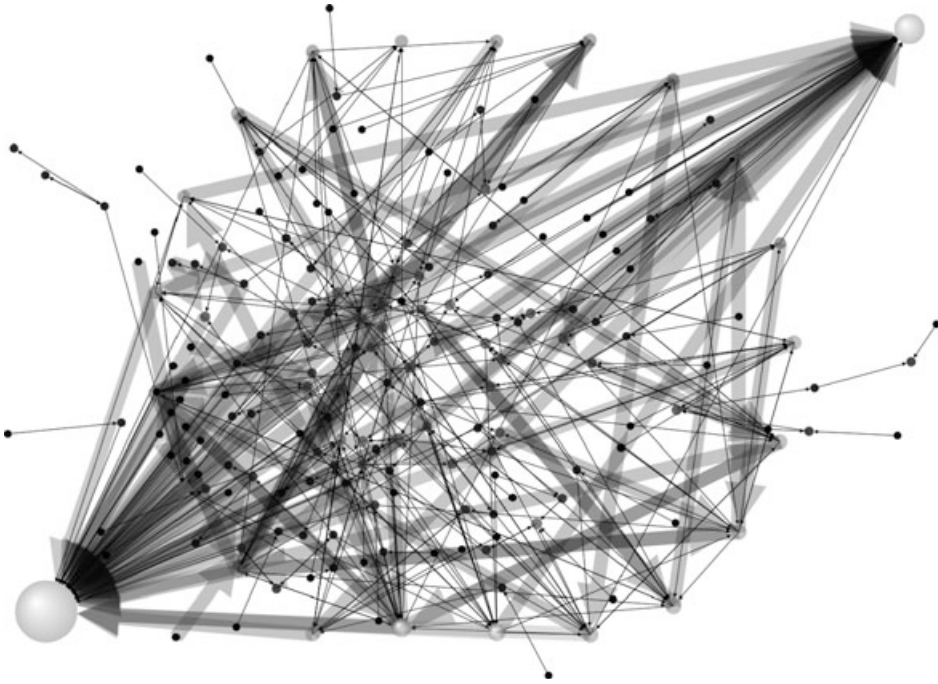


Figure 2 Two discussion catalysts in ab.politics.

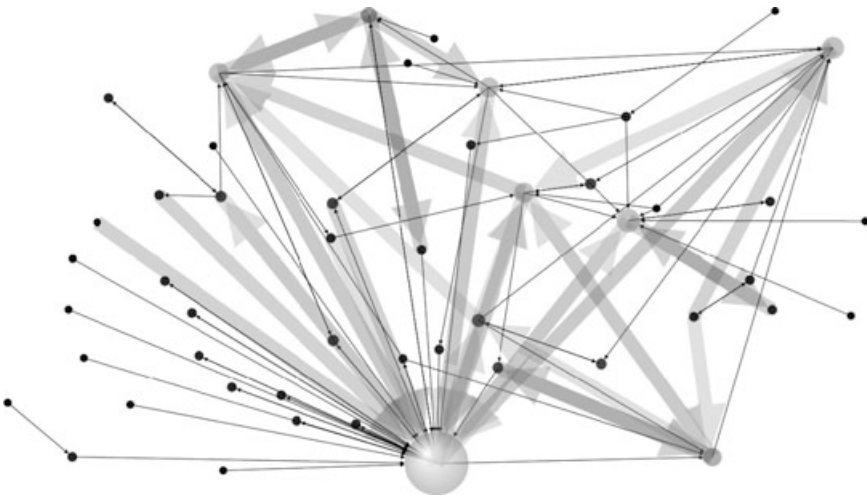


Figure 3 A discussion catalyst in alt.politics.england.misc.

attracted a significantly larger numbers of replies and repliers than any other contributor, occupying the role of discussion catalyst alone in that discussion space. In Figure 2 the authors arrayed in a circle in the center of this chart are the 10 next

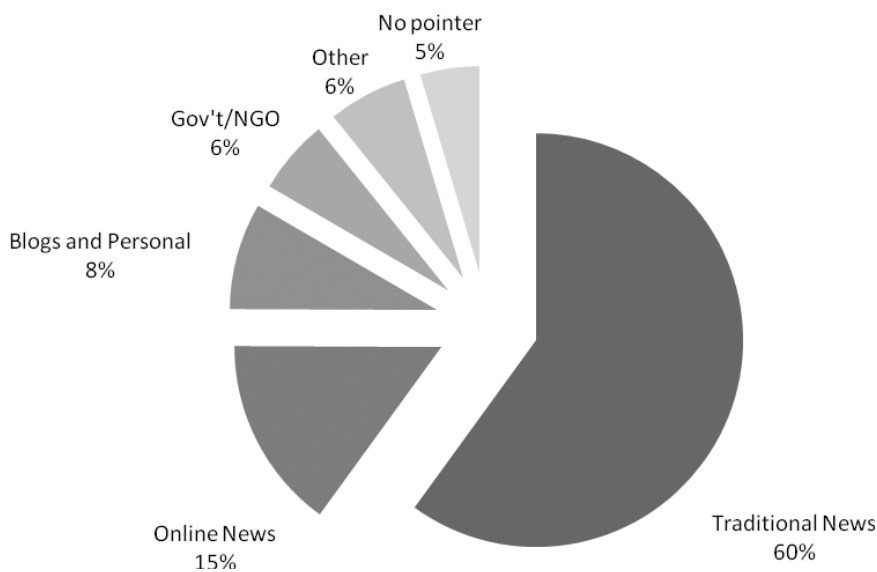


Figure 4 Major information sources in root messages (N = 325).

highest in-degree participants in the newsgroup. The resulting image illustrates the disproportionate contributions made by the few high in-degree contributors. While similar in terms of their inward connections, the discussion catalysts in figure 2 display more outward connections than displayed by the author in figure 3.

Content analysis

The content of thread starting messages is of particular interest. What common features do messages that start large threads possess? To assess the content of messages that start large threads we examined 382 messages, 10 percent of all root messages posted by discussion catalysts. Of them, 57 could not be retrieved¹, leaving 325 root messages for analysis.

Of these root messages, 95.4 per cent (310 messages) included imported content from sources on the World Wide Web as pasted raw articles or URLs and 4.6 per cent (15 messages) included only original content. Of all 325 root messages, 60 percent included linked to or pasted content from traditional media websites. The leading news organizations were Associated Press (24 times), the Washington Post (23), and the New York Times (11). Other major sources were online-only news sites, such as Salon.com (15 percent), blogs and personal websites, such as Capitol Hill Blue (8 percent), and government, such as the White House, and nonprofit organizations, such as Citizens for Legitimate Government (six percent). Figure 4 summarizes these findings.

Of the root messages that imported content from the web, 65 percent contained a brief comment (up to two sentences) by the author, such as '[t]he guy's definitely gettin' to be very unpopular' and 'stupid'. Seven per cent included both imported

content and substantial contribution (more than two sentences) by authors, while the rest had no personal comments. Of all analyzed root messages, only 12 percent included a substantial contribution by the authors that was apparently original content.

Discussion

More than half a century ago, Katz and Lazarsfeld (1955) introduced the idea that information from the world of content creators, namely mass media, flows to individuals through paths mediated by a few individuals who are highly active in consuming and amplifying messages. Despite the recent dramatic changes in communication technologies, this study illustrates how the fundamental idea of a small population wielding disproportionate influence on the flow of information still stands. Following one major criticism of Katz and Lazarsfeld's theoretical framework (Weimann, 1994), we separate the flow of information from the flow of influence and focus on the former.

This study identified a social role, the discussion catalyst, which has an important function in political discussion newsgroups. As one of a few highly replied-to participants, the discussion catalyst influences what enters the newsgroup and affects what happens within it. In the flow of information to the newsgroup, they are content importers who bring mainly news articles for discussion to their newsgroups with little or no comment. Within these newsgroups, they attract a large and disproportionate number of replies and repliers to the threads they initiate and thus amplify the discussion around topics they introduce.

Like their predecessors in the world of mass media, discussion catalysts bring information to these newsgroups from the external world of content. Interestingly, the similarity in roles across time and technology does not stop there. Most of the imported content comes from traditional news organizations. While these sources dominated, alternative sources were present, accounting for a third of all messages with pointers to third-party content.

The process of nominating topics for discussion continues to be mediated by a few gatekeepers who act as filters and amplifiers of mainstream media messages. Internet discussion spaces place fewer barriers to entry for those who wish to nominate topics for discussion among a potentially large group of participants and readers. However, while discussion catalysts can self-nominate, they cannot self-ratify. Only the behavior of other participants in the discussion space can convert an attempt at catalyzing discussion into a successful thread that attracts many repliers. Discussion catalysts attract many of the repliers in the community and start a disproportionate number of the topics that successfully attract replies.

This study also contributes to the growing body of literature on social roles, specifically social roles in computer-mediated discussion spaces. We took an approach similar to prior work to identify the unique role of discussion catalyst. In contrast to technical support groups, the influential people in political newsgroups were

those who attracted many replies rather than those who sent many replies. Political campaigns and Internet search engines might both consider these methods, potentially applied in near real-time, for identifying political topics and discussions that are actually discussed and the authors who are best at identifying those topics. Internet search engines could consider this method as a means of delivering a form of social search that constructs an alternative to page rank that implements a form of “people rank” built on the social structure of an author’s discussion network (Brin & Page, 1998).

Very large collections of records of social interactions via the Internet, such as the one used in this study, are increasingly available to researchers. We contribute a methodological approach for identifying the few threads and participants who are of great relevance for a study of discussion on the Internet, in general, and political discussion in particular. The three behavior ratio thresholds used to identify discussion catalysts allowed us to reduce more than 16,000 authors to 28 authors sharing a rare but socially critical role.

Limitations and directions for future studies

This study is an initial exploration of the patterns of discussion initiation and response. Several limitations are worth noting which point to future work. This study examined discussion activity in a limited number of forums in one form of social media (newsgroups) during a limited time scope. We specifically selected individuals who were very successful in starting new threads, however we did not compare them to other individuals who were less successful. We also do not have insights regarding the demographics of discussion catalysts.

While the scope of this study was political discussions, it provides the ground for the analysis of discussions of other topic. Expanding the scope of the analysis beyond political newsgroups would help broaden the relevance of the role of discussion catalyst. Exploring repositories of political discussion outside the Usenet is another step towards validating the presence of these roles in other computer-mediated social spaces like web boards, forums, and e-mail lists. Political discussions have particular interest for many scholars but other topical areas, like discussions of health, lifestyles, and culture, also bear investigation. Other dimensions of data could be introduced to further refine our understanding of the discussion catalyst and related roles. The network structure of each role can be subject to far more extensive analysis that would explore the centrality of the role within the discussion newsgroup’s social network. These measures may shift focus on the roles that are most closely related to discussion catalysts, like the discussion person who focuses activity on replying to threads.

This project did not apply content analysis to confirm that a reply does in fact refer meaningfully to the new topic introduced in the initial message. This is a direction future research may take. In the current work, we assume that a reply is a kind of vote to focus attention on the topics introduced in the initial message. While replies may

be a crude proxy of influence and attention, at the very least, replies add visibility to the initial message by contributing to the volume of messages associated with the starting message.

Conclusions

The Internet changes the way we access information and alters our ability to contribute to large pools of opinions and ideas. This study explored the flow of information among participants in a set of political online discussion forums to answer three questions: how is the success in starting large threaded discussions distributed; what characterizes the content of large thread starting messages; and where does the content of these messages originate? We found that information in these discussions came largely from the World Wide Web. Those who start discussions that attract many replies and repliers play a unique role in shaping the discussion. The distribution of successful topic starting practices in online political discussions is highly skewed. The content of large thread starting messages was largely drawn from major commercial sources of information and news.

Individuals who start online discussions play a key social role in mediating the flow of information. First, without regard to the potentially egalitarian nature of the Internet, it is still the case that a small minority of individuals amplifies selected elements of the flow of information to political newsgroups. Second, this study highlighted patterns of information flow between mainstream media and individuals via computer-mediated communication technologies. Interestingly, the role these individuals play on the Internet resembles the social role of the influential opinion leader (Katz & Lazarsfeld, 1955). Discussion catalysts import content from elsewhere on the web, by either linking to or pasting content. This is encouraging news, because the range of sources of information that discussion catalysts link to or copy from is broad and includes traditional news sources along with alternative sources such as blogs. Third, most of the sources "cited" by discussion catalysts came from more traditional news sources. This is less encouraging news, because alternative information sources are used far less than traditional news organizations, regardless of the wide range of sources of information now available.

What might explain the findings? The mere existence of discussion catalysts may have a range of explanations. From a network theory perspective, many large networks take a preferential attachment form (Barabási & Albert, 1999; Newman, 2001, 2003), in which a few elements in a network attract a large and disproportionate number of links. By conceptualizing a discussion as a network of participants who reply to one another, the dominance of a few participants is expected. From a psychological perspective, the richness of information may push people to reduce information overload. Jones, Ravid, & Rafaeli (2002) suggested that individuals could adopt a range of strategies to reduce the impact of information overload from Usenet newsgroups, for example, by failing to respond to certain messages or people. Replying to familiar discussion starters can be a useful strategy. The finding that

most of the information was imported from news media may be explained by the relatively early stages of the Internet. True, technology develops fast, but people's habits may not. Discussion catalysts and other newsgroup participants may simply trust the more familiar and established information sources more than the newer ones.

The Internet provides discussion platforms on almost any topic one can imagine. Political discussions were chosen deliberately for this study given the role that political discussions play in civil societies and the hope that the Internet will overcome some of the limitations of traditional and profit-driven mass media. The rapid development of new and grassroots information sources has limited effect on political discussions, as most information was imported from traditional news media organizations. Regardless of the potential of the Internet to provide a more egalitarian space for information and opinion exchange, few individuals still influence what others will discuss.

Notes

- 1 Discrepancies between Netscan and Google system's handling of events like cancelled messages account for the later system's inability to identify the content of 57 message IDs.

References

- Baker, C.E. (1989). *Human Liberty and Freedom of Speech*. New York: Oxford University Press.
- Barabaoi, A.L. & Albert, R. (1999). Emerging of scaling in random networks. *Science*, **286**, 509–512.
- BBC (2007). Yahoo 'censored' Flickr comments, May 18 2007. [online] Available at <http://news.bbc.co.uk/1/hi/technology/6665723.stm> (June 4, 2008).
- Brin, S. & Page, L. (1998). The anatomy of a large-scale hypertextual Web-search engine. *Computer Networks*, **30**, 107–117.
- Brosius, H. B. & Weimann, G. (1996). Who sets the agenda? Agenda-setting as two-step flow. *Communication Research*, **23**, 561–580.
- Bucy, EP, & Gregson KS (2001). Media participation: A legitimizing mechanism of mass democracy. *New Media & Society*, **3**(3), 357–380.
- Burkhalter, B. & Smith, M. (2004). Inhabitant's uses and reactions to Usenet social accounting data. In D. N. Snowden, E. F. Churchill & E. Frecon (Eds.), *Inhabited information spaces* (pp. 291–305). London: Springer-Verlag.
- Bagdikian, B. (2004). *The new media monopoly*. Boston: Beacon Press.
- Corrado, A. & Firestone, C. M. (eds.) (1996). *Elections in cyberspace: Toward a new era in American politics*. Washington, DC: Aspen Institute.
- Donath, J. S. (1996). Identity and deception in the virtual community. In P. Kollock & M. Smith (Eds.), *Communities in cyberspace* (pp. 29–59). London: Routledge.

- Fisher, D., Smith, M., & Welser, H.T. (2006). You are who you talk to: Detecting roles in Usenet newsgroups. *Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06)*, p. 59b.
- Gitlin, T. (1978). Media sociology: The dominant paradigm. *Theory and Society*, **6**, 205–252.
- Golder, S.A. & Donath, J. (2004). Social roles in electronic communities. Paper presented at Association of Internet Researchers IR 5.0, September 19–22, 2004, Brighton, England.
- Hauben, M. & Hauben, R. (1997). *Netcitizen*. London: Wiley.
- Haythornthwaite, C. & Hagar, C. (2005). The social world of the Web. *Annual Review of Information Science and Technology*, **39**, 311–346.
- Herring, S. C. (2004). Slouching toward the ordinary: Current trends in computer-mediated communication. *New Media & Society*, **6**(1), 26–36.
- Kahn, R. & Kellner, D. (2004). New media and internet activism: From the ‘Battle of Seattle’ to Blogging’. *New Media & Society*, **6**, 87–95.
- Katz, E. (1987). Communications research since Lazarsfeld. *Public Opinion Quarterly*, **51**, s25–s45.
- Katz, E. & P. Lazarsfeld (1955). *Personal influence*. New York: The Free Press.
- Levine, P. (2000). The Internet and civil society. *Philosophy and Public Policy*, **20**(4), 1–8.
- Lazarsfeld, P., Berelson, B., & Gaudet, H. (1948). *The people’s choice*. New York: Columbia University Press.
- McMillan, R. (2008). Bill Gates: Internet censorship won’t work. *IDG News Service/San Francisco Bureau*. Retrieved July 12, 2008, from http://www.nytimes.com/idg/IDG_002570DE00740E18882573F50010C487.html.
- Mehra, B., Merkel, C., & Peterson Bishop, A. (2004). The Internet for empowerment of minority and marginal users. *New media and society*, **6**(6), 781–802.
- Newman, M.E.J. (2001). Clustering and preferential attachment in growing networks. *Physical Review E*, **64**(2 Pt 2), 025102.
- Newman, M.E.J. (2003). The structure and function of complex networks. *SIAM Review*, **45**(2), 167–256.
- Papacharissi, Z. (2002). The virtual sphere: the internet as a public sphere. *New Media and Society*, **4**(1), 9–27.
- Picard, R.D. (1985). *The press and the decline of democracy: The democratic socialist response in public policy*. Westport, Connecticut: Greenwood Press.
- Rheingold, H. (1993). *The virtual community: Homesteading on the electronic frontier*. Reading, MA: Addison-Wesley.
- Siebert, F.S., Peterson, T., & Schramm, W. (1963). *Four theories of the press*. Freeport, NY: Books of Libraries Press.
- Southwell, B. G. & Yzer, M. C. (2007). The roles of interpersonal communication in mass media campaigns. *Communication Yearbook*, **31**, 420–462.
- Turner, T. C., Smith M., Fisher, D., & Welser H. T. (2005). Picturing Usenet: Mapping computer-mediated collective action. *Journal of Computer Mediated Communication*, **10**(4), Retrieved June 4, 2008, from <http://jcmc.indiana.edu/vol10/issue4/turner.html>

- Turner, T.C. and Fisher, K.E. (2006). The impact of social types within information communities: Findings from technical newsgroups. *Proceedings of the 39th Annual Hawaii International Conference on System Sciences*, **6**, 135b.
- Van Den Ban, A. W. (1964). A revision of the two step flow of communication hypothesis. *Gazette*, **10**, 237–249.
- Viegas, F.B. & Smith, M.A. (2004). *Newsgroup crowds and author lines: Visualizing the activity of individuals in conversational cyberspaces*. Big Island, HI: IEEE.
- Weimann, G. (1994). *The influentials: People who influence people*. New York: State University of New York Press.
- Welser, H.T., Gleave, E., Fisher, D., & Smith, M.A. (2007). Visualizing the signatures of social roles in online discussion groups. *Journal of Social Structure*, **8**. Retrieved June 4, 2008, from <http://www.cmu.edu/joss/content/articles/volume8/Welser/>
- Wikipedia (2007). *Usenet*. Retrieved June 4, 2008, from <http://en.wikipedia.org/wiki/Usenet>

About the Authors

Itai Himelboim is an assistant professor in the Telecommunications Department, Grady College, at the University of Georgia. His research focuses on online social networks at the individual and institutional levels.

Address: 101L Journalism Building, University of Georgia, Athens, GA, 30602

Eric Gleave is graduate student in the University of Washington Department of Sociology. His research explores the practical methodological issues involved in studying social processes large online communities, especially network analysis.

Marc Smith is Chief Social Scientist at Telligent Systems, a social media platforms and services company in Dallas, Texas. His research focuses on the visualization of social media networks and computer-mediated collective action.

Email: 17950 Preston Rd., Suite 310, Dallas, Texas 75252

This project was funded in part by a research grant from Microsoft Research.